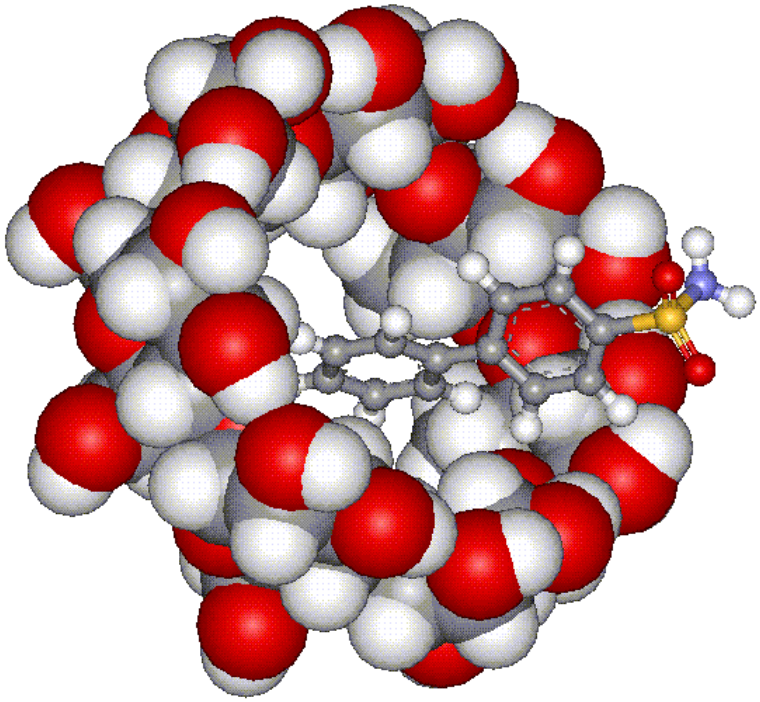
BOSS, Version 4.6

Biochemical and Organic Simulation System

User's Manual

for *UNIX, Linux,* and *Windows*



William L. Jorgensen
Department of Chemistry, Yale University
P. O. Box 208107
New Haven, Connecticut 06520-8107

Phone: (203) 432-6278
Fax: (203) 432-6299
Internet: william.jorgensen@yale.edu

August 2004

Contents **Page**

	Introduction		**4**
1	**Can't Wait to Start – Use the x Scripts**		**6**
2	**Statistical Mechanics Simulation – Theory**		**6**
3	**Energy and Free Energy Evaluation**		**7**
4	**New Features**		**9**
5	**Operating Systems**		**14**
6	**Installation**		**14**
7	**Files**		**14**
8	**Command or bat File Input**		**16**
9	**Parameter File Input**		**18**
10	**Z-matrix File Input**		**27**
	10.1	**Atom Input**	**31**
	10.2	**Geometry Variations**	**31**
	10.3	**Bond Length Variations**	**31**
	10.4	**Additional Bonds**	**31**
	10.5	**Harmonic Restraints**	**32**
	10.6	**Bond Angle Variations**	**32**
	10.7	**Additional Bond Angles**	**33**
	10.8	**Dihedral Angle Variations**	**33**
	10.9	**Additional Dihedral Angles**	**34**
	10.10	**Domain Definitions**	**35**
	10.11	**Conformational Searching**	**35**
	10.12	**Sample Z-matrix**	**36**
	10.13	**Z-matrix Input for Custom Solvents**	**36**
11	**PDB Input**		**38**
12	**Coordinate Input in Mind Format and Reaction Path Following**		**39**
13	**Pure Liquid Simulations**		**40**
14	**Cluster Simulations**		**40**

15 Solventless and Molecular Mechanics Calculations 41

15.1 Continuum Simulations 41
15.2 Energy Minimizations 42
15.3 Dihedral Angle Driving 43
15.4 Potential Surface Scanning 43
15.5 Normal Coordinate Analysis 44
15.6 Conformational Searching 46

16 Available Solvent Boxes 49

17 Ewald Summation for Long-Range Electrostatics 51

18 Test Jobs 52

19 Output 58

19.1 Some Variable Definitions 59

20 Contents of the Distribution Files 60

21 Appendix – No. of Molecules for Binary Solvent Mixtures 62

22 Appendix - OPLS All-Atom Parameters 63

Biochemical and Organic Simulation System

Introduction

The **BOSS** program performs (a) **Monte Carlo (MC) statistical mechanics simulations** for solutions of zero to 25 solute molecules in a periodic solvent box, in a solvent cluster, or in a dielectric continuum including the gas phase, and (b) **standard energy minimizations, normal mode analysis, and conformational searching**. The energetics are represented with the **OPLS force fields** or the **AM1, PM3, PDDG/PM3 and PDDG/MNDO** semiempirical molecular orbital methods. Besides the quantitative structural and energetic results, other quantities are computed including dipole moments, CM1 and CM3 atomic charges, solvent accessible surface areas, and hydrogen-bond counts that are valuable for many applications such as **structure-activity relations**. Some important features are:

1. For the Monte Carlo simulations, the **NPT ensemble** is employed, though NVT simulations can also be performed by not allowing volume changes.

2. **Preferential sampling** is used whereby the solutes and nearby solvent molecules are moved more frequently than distant solvent molecules. The biasing is based on $1/(r^2 + wkc)$ where *wkc* is a constant. See *J. PhBe fhem.c*

(((-mTf12 0 0 12274..2092 02.59952 TmgTf53 1 Tf -0.0029 Tc 0.0029 Tw 12 0 0 12237..2092 02.59952 TmL Am. Chem
d by

- *J. Am. Chem. Soc.* **114**, 7535 (1992)
- *J. Am. Chem. Soc.* **116**, 2199 (1994)

- host/guest binding
  - *J. Org. Chem.* **64**, 7439 (1999)
  - *J. Am. Chem. Soc.* **120**, 5104 (1998)
  - *Proc. Natl. Acad. Sci.* USA **90**, 1194 (1993)
- clusters
  - *J. Chem. Phys.* **99**, 4233 (1993)
- for reviews, see:
  - *Acc. Chem. Res.* **22**, 184 (1989)
  - *Encyclopedia Comp. Chem.* **2**, 1061 (1998)
- QM/MM calculations
  - *J. Phys. Chem. B* **102**, 1787 (1998)
  - *J. Phys. Chem. B* **106**, 8078 (2002)
  - *J. Am. Chem. Soc.* **126**, 9054 (2004)

5. The **OPLS** united-atom (*J. Am. Chem. Soc.* **110**, 1657 (1988)) and all-atom (*J. Am. Chem. Soc.* **118**, 11225 (1996)) potential functions are used with additional references given in the parameter files. Coverage includes numerous organic functional groups, all common peptide residues, and nucleotides. More parameters can easily be added to the parameter file. **Unpublished parameters** including for heterocycles, sulfonamides, and other pharmacologically important groups are also provided to BOSS users.

   Alternatively, the energetics of the solute(s) can be described with the **AM1, PM3, PDDG/PM3 or PDDG/MNDO** semiempirical molecular orbital methods. Any solvent-solvent interactions are described with the OPLS potentials for mixed **QM/MM calculations**. The solute-solvent interactions are described with the force field using standard OPLS-AA Lennard-Jones parameters for the solute(s) and solute charges obtained from the CM1 or CM3 procedures or as Mulliken charges.

6. Dihedral angles, bond angles and bond lengths in the solutes can be sampled during the simulations, *e.g.*, a pmf could be determined for the separation of two flexible molecules or binding to a flexible host could be studied. The solvent molecules can also be **fully flexible**.

7. **Harmonic restraints** can be included between pairs of solute atoms or between an atom and a fixed point. This facilitates some free energy calculations for host/guest systems; it can also be used for restrained optimizations and NMR structure refinement.

8. Coordinate files may be **output in PDB or MDL mol format** for easy display with standard molecular graphics programs.

1 Can't Wait to Start – Use the x Scripts

If you want to use BOSS immediately after following the installation instructions, many scripts are provided to execute standard jobs. All that is needed is a coordinate file (BOSS Z-matrix, .mol. .mol2, or .pdb). Hundreds of BOSS Z-matrix files (.z) are provided in the *aazmat* and *drugs* directories. The scripts are in the *scripts* directory and most are also duplicated in the *aazmat* and *drugs* directories. The scripts are described in *boss/scripts/00README* (*c:\boss\scripts\00README.doc* under *Windows*) and cover many standard jobs including energy minimizations with the OPLS force field or with AM1 or PM3 calculations, conformational searching, and conversion of external coordinate files (.mol, .mol2, .pdb) to BOSS Z-matrix files (.z).

Some examples follow. Output will be in the files *out*, *sum*, and *plt.pdb*. The output structure in *plt.pdb* can be displayed with many graphics programs such as RasMol, WebLab Viewer, and Midas.

(1) Perform an energy minimization for butane with the OPLS-AA force field,
cd ~/boss/aazmat (for Windows: cd c:\boss\aazmat)
xOPT butane

(2) Perform a PDDG/PM3 optimization for AZT,
cd ~/boss/drugs (for Windows: cd c:\boss\drugs)
xPDGOPT azt

(3) Convert a Tripos .mol2 file for Zoloft to a BOSS Z-matrix and optimize the structure. The OPLS-AA force field is used except the partial charges are obtained from an AM1/CM1A calculation. The file *zoloft.z* is created.
cd ~/boss/aazmat
xMOL2Z zoloft

(4) Perform a conformational search for alanine dipeptide using 100 random starting structures and include GB/SA solvation. Output is in *alanin.cs.out*, and *alanin.cs.sum*, and the structures of the conformers are output in MDL .mol format in the file *alanin.cs.mol*
cd ~/boss/aazmat
xCSGB100 alanin

2 Statistical Mechanics Simulations—Theory

Statistical mechanics simulations of many particle (*e.g.*, condensed phase) systems are commonly done by one of two main approaches, Monte Carlo (MC) and molecular dynamics (MD). The basic principles are similar, but the methodology is fundamentally different. Common to both is the use of a classical energy function to calculate the total potential energy of the system as a function of the bonds, bond angles, torsions, and interatomic interactions. The expression used in BOSS is typical for molecular mechanics (MM) force fields:

$$E_{total} = \sum_{bonds} K_r (r - r_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{torsions,n} \frac{V_n}{2}[1 \pm \cos(n\phi)] + \sum_{non-bonded} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}} \right]$$

In a QM/MM calculation, the intrasolute and intersolute interaction terms are replaced by the total QM energy. Once the potential energy of the system is known, a MC simulation generates a new configuration by a set of random motions. The difference in energy between the new and the old configuration is used as a selection criterion by the Metropolis algorithm, which enforces a correct Boltzmann distribution of energies for the system at the desired temperature, and the procedure is then iteratively repeated. The ensemble of configurations generated in this fashion can be used to compute the structure and thermodynamic properties of the system at the specified conditions. Differences in free energies between two similar states of the system can be calculated from this ensemble through the use of statistical perturbation theory. In a MD simulation, on the other hand, the forces from the energy components in the three Cartesian directions acting on each atom are evaluated. The accelerations are calculated from these forces, and a trajectory is generated by the repeated numerical integration of the equations of motion over a period of time. This trajectory constitutes an ensemble of configurations that is formally equivalent to an ensemble generated via a MC simulation and can be used in the same fashion.

This difference in methodology, although it may not be obvious at first look, has significant implications in the use and application of the methods. The major difference from an algorithmic point of view is that MC sampling can readily use internal coordinates, while MD works with independent atomic motions in Cartesian space. The consequences of this difference are many; in MD it is easy to treat large, flexible molecules since any coupling between different degrees of freedom is automatically handled by the algorithm. On the other hand, the integration time step has to be very small in order to avoid numerical instability and imprecision. There is also a certain inefficiency in the sampling since motions that go uphill in energy are usually reversed by opposing restoring forces. Also, the algorithm makes the selective freezing of unimportant degrees of freedom cumbersome. In a MC calculation, on the other hand, the freezing of selected internal coordinates is easily accomplished by simply not allowing variations in the specified bonds, angles, or dihedrals; however, sampling for large flexible molecules can be problematic since degrees of freedom are treated, for the most part, as being completely uncoupled. An additional advantage of MC is that it is possible to sample large regions of configurational space by stepping over potential energy barriers. A paper comparing the efficiencies of MD and MC in simulations of liquid hexane has appeared: W. L. Jorgensen & J. Tirado-Rives, *J. Phys. Chem.* **100**, 14508 (1996). In this case, MC was found to be 2-4 times more efficient for the conformational sampling.

Summaries on the theory and computations can be found in the following paper along with references to more detailed presentations: W. L. Jorgensen, *J. Phys. Chem.* **87**, 5304 (1983). See also: M. P. Allen and D. J. Tildesley, Computer Simulations of Liquids; Oxford U. Press: London, 1987.

3 Energy and Free Energy Evaluation

In BOSS, the total MM potential energy of a system is given by eq 1

$$E = ESS + ESX + EXX + EBND + EBC + EANG + EDIH + ENB + ECUT + ESINT \quad (1)$$

where

ESS = the solvent-solvent energy,

ESX = the solvent-solute energy,

EXX = the solute-solute (intersolute) energy,

EBND = the bond stretching energy for the solutes,

EBC = the energy for the harmonic restraints,

EANG = the angle bending energy for the solutes,

EDIH = the torsional energy for the solutes,

ENB = the >1,3 intramolecular non-bonded energy for the solutes, and

ECUT = the cutoff correction for the Lennard-Jones interactions neglected beyond the cutoff for non-aqueous solvents,

ESINT = the intramolecular energy for flexible solvent molecules,
= ESBND + ESANG + ESDIH + ESNB

For MM calculations, the Coulombic energy can be evaluated with spherical cutoffs or by the Ewald method. In a QM/MM calculation, EXX, EBND, EANG, EDIH, and ENB are replaced by the total quantum mechanical energy, EQM.

Free energy changes, if desired, are computed with statistical perturbation theory in a windowing format with double-wide sampling. ΔG (Gibbs) and ΔA (Helmholtz) are computed for NPT and NVT simulations, respectively. The expression for ΔG is in eq 2 where the perturbation is from the reference system *i* to the perturbed system *j*. The mutation of A to B involves a scaling where a coupling parameter λ is used such that for $\lambda = 0$ the system is A and

$$\Delta G = G_j - G_i = -kT \ln \langle \exp(-(E_j - E_i)/kT) \rangle_i \qquad (2)$$

for $\lambda = 1$ it is B. Then for any geometrical variable or potential function parameters (q, σ, ε) Y, eq 3 is used to scale the system from A to B. Thus, the simulation is run at a reference

$$Y_i = \lambda_i Y_B + (1 - \lambda_i) Y_A \qquad (3)$$

value λ_i corresponding to *i* in eq 2. The program perturbs the system to two other values of λ corresponding to j and k, where typically $j < i < k$. This computation of two free energy increments in one simulation has been termed "double-wide sampling" (*J. Chem. Phys.* **83**, 3050 (1985)). The number of free energy increments that need to be computed depends on the details of the overall perturbation, the solvent and T, P conditions. Many examples are available in the references above and test jobs. For the conversion of ethane to methanol in water, 5 increments were sufficient to give excellent precision (*J. Chem. Phys.* **83,** 3050 (1985)), while the annihilation of adenine or guanine in chloroform required 30-40 increments (*Tetrahedron*, **47**, 2491 (1991)) necessitating 15-20 simulations with double-wide sampling.

The perturbations can also involve a reaction coordinate such as a dihedral angle or intermolecular distance. This is easily specified in the Z-matrix file by indicating the beginning and ending values of the variable corresponding to $\lambda = 0$ and 1. In this case there might be no change in the atom types (and potential function parameters). In BOSS, the three λ values are referred to as RC0 for the reference solute and RC1 and RC2 for the perturbed solutes (Section 8).

4 New Features in BOSS 4.6:

1. A GB/SA hydration model has been implemented for optimizations, single-point calculations, dihedral driving, conformational searching, and Monte Carlo simulations; FEP-GB/SA calculations have not been implemented yet. The GB/SA treatment is based on the model described in Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C., *J. Phys. Chem. A*, **101**, 3005-3014 (1997). A key difference is OPLS-AA sigmas divided by two ($\sigma/2$) are used as the van der Waals radii, R_{vdW}, whereas Still et al. used OPLS-UA sigmas. All hydrogens are ignored for the nonbonded term in eq 5a of their paper and only aliphatic hydrogens are ignored for the stretch and bend terms. Charges on aliphatic hydrogens are collapsed onto their adjacent carbons. The results for 75 molecules including the 35 molecules in Table 3 of their paper yield an average error of 0.61 kcal/mol. The ionic strength of the solution can be specified on line 2 of the par file; see STREN on p 18. See the preprint gbsa.pdf in the *bossman* directory for more information.

   Scripts that use GB/SA have been added including xSPGB, xOPTGB, xOPTFGB (preferred for optimizations), xCSGB100, and xMCGB. See the updated optimize, consearch, scan, and dihdrive test jobs and the new MCGBSA test job.

2. The parameterization of the PDDG/PM3 and PDDG/MNDO methods has been extended to the halogens (F, Cl, Br, I) and most recently to sulfur: Tubert-Brohman, I.; Guimaraes, C. R. W.; Repasky, M. P.; Jorgensen, W. L., *J. Comput. Chem.*, **25**, 138-150 (2004).

3. Computed atomic charges from the quantum methods (CM1, CM3, Mulliken) for equivalent, monovalent atoms are now symmetrized. Thus, for example, in actetate ion the three methyl hydrogens will have the same charge as will the two oxygens. Similarly, the two oxygens of sulfones or nitro compounds and the fluorines in a CF_3 group will have the same charge. Formally and formerly, in the absence of the appropriate symmetry, charges for such atoms are computed to be different by the quantum methods. This change simplifies the use of the quantum charges with force field calculations; it improves, for example, conformational searching by avoiding false multiple minima that would arise from the former charge asymmetry.

4. Similarly, computed CM1, CM3 and Mulliken charges for equivalent atoms in mono- and disubstituted phenyl rings are now symmetrized. E.g., for phenol, the ortho carbons now have the same charge as do the ortho hydrogens, etc. For *m*-cresol, the charges for C4 and C6 are the same as are the charges for their attached hydrogens.

5. In an expanded study of FEP computation of free energies of hydration using the CM1 and CM3 procedures, we have found the lowest average errors using CM1A charges scaled by 1.14 for neutral molecules. Charges for ions are not scaled. For 25 diverse organic molecules, the average unsigned error in absolute free energies of hydration is 1.0 kcal/mol with the 1.14*CM1A charges and OPLS-AA otherwise. Ref: Udier-Blagović, M. et al. *J. Comput. Chem.*, **25**, 0000-0000 (2004).

6. The charge symmetrization has been applied to all Z-matrices for molecules in the *boss/drugs, boss/antib, and boss/drugHIV* directories. These Z-matrices now use the CM1A charges, scaled by 1.14 for neutral molecules. Improvements which were made to the atom-typing for five-membered heteroaryl rings are also reflected in the new Z-matrices.

7. For conformational search, the new utilities xCSZ and xGETE have been added. xCSZ separates the concatenated *molecule.cs.z* file into separate files, *cs001.z*, *cs002.z*, etc. xGETE is used to extract the Total Energy that is on line 1 of Z-matrices output by BOSS. For example, one could run a conformational search using the OPLS-AA force field, then optimize each output conformer with PDDG/PM3, and collect the resultant energies in the *energy.CSV* file, which could be merged with *molecule.cs.CSV* –
xCS200 *molecule*
xCSZ (input *molecule.cs.z*)
ls -1 cs*.z >>*csfiles* (lists created file names in *csfiles*)
xMANY *csfiles* xPDGOPT (does PDG optimizations & creates *cs001.pdg.z*, etc.)
ls -1 cs*pdg.z >>*pdgfiles*
xMANY *pdgfiles* xGETE (adds energies to *energy.CSV*)
then merge *molecule.cs.CSV* and *energy.CSV*; the resultant CSV file has the OPLS-AA results for all conformers and the PDDG/PM3 optimized energies.

7. See NONEBN on page 24. Its use can guarantee a correct covalent neighbors list.

New features in BOSS 4.5 were:

1. PDDG/PM3 and PDDG/MNDO calculations have been implemented for elements C, H, N, and O: M. P. Repasky, J. Chandrasekhar, and W. L. Jorgensen, *J. Comput. Chem.*, **23**, 1601-1622 (2002).

2. Ewald summation can be used to evaluate the Coulombic energy for MC simulations with the OPLS force fields. See Section 17.

3. The CM3 charge models have been implemented in addition to CM1 for AM1 and PM3 wavefunctions: J. D. Thompson, C. J. Cramer, and D. G. Truhlar, *J. Comput. Chem.*, **24**, 1291-1304 (2003).

4. New scripts continue to be added. Check the 00*README* file in the *scripts* (formerly *cmdpar*) directory. E.g., see *xPDGOPT*, *xFEPAQ*, *xFEPGAS*, and *xMCGAS*. All scripts have been modified to use *scripts* rather than *cmdpar*.

5. Bond orders are now included in output MDL *.mol* files for better display with WebLabViewer.

6. Improvements have been made for MC simulations of pure liquids – please see *notes/pureliquids*. Also see the new *scan* and *SN2rxn* test jobs for transition structure searching and automated determination of reaction profiles in solution.

7. For Monte Carlo conformational search, a summary of the results is now written to a CSV file – try one of the xCS scripts. Other improvements have been made for both MC and stochastic conformational search.

New features in BOSS 4.3 were:

1. The top directory for BOSS is now called *boss* and the executable is BOSS for Unix/linux and boss.exe for Windows. The test jobs directory is now *bosstst*. All command files and scripts have been changed accordingly.

2. In order to expand the OPLS-AA atom types to >999 possibilities, the atom type entry in the oplsaa.par and oplsua.par files has changed from I3 to I4 format. One must use these new, up-to-date par files or ones in the same format with BOSS 4.3.

3. Treatment of large molecules has expanded. (a) For energy minimizations and conformational search, up to 450 variable degrees of freedom can be optimized. (b) For AM1 and PM3, up to 100 non-hydrogen and 150 hydrogen atoms are allowed.

4. Many new molecular structures have been added in the *aazmat* and *drugs* directories. Antibiotics are in the *antib* directory and many anti-HIV drugs are in *drugHIV*.

5. The OPLS-AA parameter files, oplsaa.par and oplsaa.sb, have been updated with many new entries including the parameters recently published for esters, nitriles, nitro compounds, and perfluoroalkanes:
Gas-Phase and Liquid-State Properties of Esters, Nitriles, and Nitro Compounds with the OPLS-AA Force Field. M. L. P. Price, D. Ostrovsky, and W. L. Jorgensen, *J. Comput. Chem.* **22**, 1340-1352 (2001).
Perfluoroalkanes: Conformational Analysis and Liquid-State Properties from Ab Initio and Monte Carlo Calculations. E. K. Watkins and W. L. Jorgensen, *J. Phys. Chem. A* **105**, 4118-4125 (2001).

New features in BOSS 4.2 were:

1. The TIP5P water model has been added as a stored solvent (*J. Chem. Phys.* **112**, 8910 (2000)).

2. For linear response calculations, the numbers of carboxylic acids, amides, nitro groups, unconjugated amines, and rotatable bonds in the solutes are perceived and output at the end of the sum file. They are read by the PROPAA and PROPCMP utilities for fully automated prediction of pharmaceutically important ADME properties including octanol/water log P, Caco-2 cell permeability, and aqueous solubility. The SASA decomposition now contains four terms – hydrophobic, hydrophilic, carbon π, and weakly polar (halogens, S and P).

3. For custom solvents, the solvent-solvent and solute-solvent cutoffs can include distances to up to five solvent atoms declared in the ICUTAS array in the par file. This is for large solvent molecules. Formerly, only one distance to NCENTS was used.

4. Many scripts have been provided in the scripts directory for facile execution of standard BOSS jobs. E.g., the command "xOPT molecule" executes an energy minimization with the OPLS-AA force field for the molecule with Z-matrix file, molecule.z; "xPM3SP molecule" executes a PM3 single-point calculation, etc.

5. To facilitate the study of reaction paths, improvement was made to promote SCF convergence for AM1 and PM3 calculations. If the initial guess based on the density matrix does not lead to convergence, a new initial guess is made from diagonalization of the core Hamiltonian. Subsequent calculations will attempt both initial guesses, if necessary. No convergence problems have been encountered with this change.

6. Stochastic conformational search has been implemented (M. Saunders, *J. Am. Chem. Soc.* **109**, 3150 (1987)). This is particularly valuable for rings. Also, Monte Carlo conformational search now requires no special input in the Z-matrix. See Section 10.11.

7. Most command files have been modified so that the par files no longer contain the non-bonded and torsional parameters. Instead the latest version of the oplsaa.par file is appended to the top part of the par file via commands in the command file.

8. The format for the torsion parameters in the par file has changed. The V0 parameter has been removed and four-fold torsions with V4 parameters are now allowed. The phase angles have also been removed. Old par files will still be processed properly.

9. The execution time for geometry optimizations including conformational searching has been decreased significantly by use of Cartesian derivatives. One can perform optimizations with symmetry constraints for bond lengths, bond angles, and dihedral angles. See pages 26-27.

New features in BOSS 4.1 were:

1. Additional dihedral angles can now be determined automatically - Section 10.9. The program automatically assigns their types.

2. The types and ranges for variable dihedral angles can also be assigned automatically - Section 10.8. These two additions vastly simplify the BOSS Z-matrix input.

3. A utility program, *autozmat*, is provided that interconverts many structure file types, e.g., PDB, Sybyl mol2, MDL mol, Gaussian Z-matrix, and BOSS Z-matrix. One can now go from a Sybyl mol2 file to a fully-flexible BOSS Z-matrix in a second. For more information enter *autozmat -h*, when in the boss directory. Also see Section 18.1.

4. A flat-bottomed harmonic potential has been added for harmonic restraints - Section 10.5. Dihedral angles can also be harmonically restrained - Section 10.9. These additions are useful for refinement of NMR structures using NOE and coupling constant data.

1. The solvent accessible surface area is now decomposed into hydrophobic, aromatic, and hydrophilic components. These are averaged over MC runs and tabulated with standard deviations.

2. The optional drugs subdirectory contains optimized BOSS Z-matrices for ca. 100 drugs.

New features in BOSS 4.0 were:

1. The ability to perform QM/MM calculations has been added (see QMNAME - Section 9). The energetics for the solutes can be described with AM1 or PM3. The solute charges come from Mulliken populations or via the CM1A or CM1P procedures of J. W. Storer, D. J. Giesen, C. J. Cramer, and D. G. Truhlar, *J. Comput.-Aided Mol. Des.* **9**, 87 (1995). Some results with the AM1/CM1A combination were described in *J. Phys. Chem. B* **102**, 1787 (1998). For example, this addition allows studies of solutes with full flexibility in solution with no need for OPLS parameters for the solutes. It also

permits reaction paths to be followed in solution with evaluation of the energetics for the solutes on-the-fly (see test job rxnpmf).

2. Additional bond angles can now be determined automatically by the program. They no longer have to be declared explicitly - Section 10.7.

3. The test jobs have been extensively redone to be both more pedagogical and representative of standard calculations executed with BOSS - Section 18.

Some New Features in Earlier Versions of BOSS were:

The *solvent accessible surface area and volume* of the solutes are automatically computed. They are determined for the solute structures and printed out at the beginning and end of a run. The solvent probe radius is input in the par file or default values are used. The atomic radii for the solute atoms are calculated from the OPLS σ non-bonded parameters; radius = $2^{1/6}\sigma/2$.

Multiple independent solute or solvent molecules can be used to search for optimal complexes. This can be useful in designing receptors for a guest molecule. (See NOXX and NOSS - Section 9.)

Heating of solutes at a different temperature than the solvent has been implemented. This "*local heating*" feature may be useful for equilibrations. (Section 9.)

Interactions beyond the cutoff can be handled with a *reaction field*. (See DIELRF - Section 9).

An atom can be restrained to a point in space - see Section 10.5.

Gas-phase optimizations can be performed using the Hooke-Jeeves, BFGS, conjugate gradient, and simulated annealing methods in addition to the existing Simplex, Powell and Fletcher-Powell methods.

Conformational searching can be carried out using random variation of internal coordinates (Section 10.11). Test job consearch illustrates a conformational search.

A utility program, pargen, has been provided to create interactively the top portion of the parameter files, which contain variables for the requested type of calculation (Section 9).

Any solvent can now be used or MC simulations performed for any pure liquid by specifying a Z-matrix for a solvent molecule. The custom solvent can be treated as a rigid molecule or have any designated internal degrees of freedom sampled (Section 13). Equilibrated boxes for several custom solvents are provided.

The vibrational frequencies for the solute(s) in the gas phase can now be computed from a normal mode analysis (Section 15.5).

The number of possible solutes has expanded from 2 to 25. Solute entries are now separated in the Z-matrix by "TERZ" lines.

Binary solvent mixtures of any composition can now be used.

5 Operating Systems

BOSS is normally distributed in *Linux* or *Windows* versions. The *Linux* executables were obtained with the Intel FORTRAN compiler on a 2.4 GHz Xeon running Red Hat release 8.0; they should work on any Linux system with a 2.0 or newer kernel. The *Windows* executable has been obtained with the Intel FORTRAN compiler, version 7.0; this yields ca. 10% execution speed gains over the Portland compiler in our hands.

IRIX executables are included in the *boss/IRIX* directory with the *Linux* version of BOSS. If you are running on SGI computers, see the README file in this directory.

The *Windows* version of BOSS runs under *Windows9x, WindowsNT, Windows2000, and WindowsXP* for PC's. A letter to the editor of the *Journal of Computational Chemistry* reporting relative timings for UNIX workstations and Pentium-based PC's is in the August 1996 issue (Vol. 17, p1385-1386). BOSS was demonstrated to execute at the same speed on a SGI R5000 workstation or a 200 MHz Pentium Pro. The *Windows* version provides a powerful molecular modeling system that can also be executed on laptop computers. Display of PDB files of structures output by BOSS can be performed well on PC's using free software such as *RasMol* by Roger Sayle at Glaxo-Wellcome, *WebLabViewer* from Accelrys, and *Chimera* from UCSF.

6 Installation

The BOSS files are normally provided by ftp or on a CD. Instructions are given for installation of the files in each case. For UNIX/Linux the path to the top BOSS directory needs to be specified in the .cshrc file with a command like
setenv BOSSdir ~/boss

The key files are placed in the BOSSdir disk directory, normally called *boss*. The test jobs are placed in separate subdirectories of the *bosstst* directory. There are README files that contain useful information that should be read. If the source code has been provided, it is in the *source* subdirectory. It can be compiled and linked to yield the executable file called BOSS.

7 Files

The following files are used by the program. The logical units are assigned in the BLOCK DATA routine and stored in COMMON/IO.

Unit	**Suffix**	**Description**
IRD	CMD	The "command" file for executing the program.
ILST	OT	Printed output file (the "out file").
IDSKI	IN	The "in file" contains the coordinates and other data needed to restart the simulation.
IDSKO	UP	The "up file" contains the results for each run that are needed to compute the global averages.

IDSKS	SV	The "save file" is used to store all coordinates periodically for later analysis. The period is specified in the parameter file.
IDSKA	AV	The "av file" contains the averages for distribution functions that may be plotted.
IZM	ZMAT	Contains the Z-matrix input for the solutes.
IZMS	ZMAT	Contains the optional Z-matrix input for a custom ("other") solvent.
IPAR	PAR	The "parameter" file contains the potential function parameters and variables for the present simulation.
IBAPAR	oplsua.sb oplsaa.sb	Contain bond stretching and angle bending parameters.
IWAT	watbox	Contains water boxes for initial system setup.
INAQ1	org1box	Contains solvent boxes for non-aqueous solvents.
INAQ2	org2box	Additional solvent boxes for non-aqueous solvents.
IPLT	PLT	The "plt file" receives output coordinates in the PDB, Mind, or MDL mol format.
ISUM	SUM	Short version of the printed output file goes to the "sum file".

Most of these files are created automatically by the program. The only user supplied files are the CMD or bat, PAR and ZMAT files. watbox, org1box, org2box, oplsaa.sb and oplsua.sb must also be in the user's disk area or in a general access area such as /usr/local/lib/. The IN file is typically empty for an initial system setup. The input for the CMD or bat, PAR and ZMAT files is described in the following sections.

The MC simulations are carried out in a series of runs, each consisting of *ca*. 200K to 1000K configurations. A new configuration is generated by randomly moving a solute or solvent molecule or by changing the volume of the system by scaling the coordinates of all molecules. The "in file" is rewritten at the end of each run. The "up" and "save" files grow during the simulation, and the latter can get quite large. If the "up file" is destroyed accidentally, all averaging information on the simulation is lost.

8 Command or bat File Input

If you are just using the x scripts, you can skip this section.

A *UNIX* command file or *Windows* bat file for BOSS execution begins with a header that can be updated. See, for example, the test jobs such as *bosstst/dihpmf*. A series of 1 to 20 calculations is typically executed with one job submission. Each begins with the file assignments. Often, the suffixes for the file names are not changed, though individuals may have their own preferences for designating file names. The program is executed by the *UNIX* command, ***e.g.***,

```
time ~/boss/BOSS 111 500000 10.0 0.0 20.0
```

or the *Windows* command,

```
c:\boss\boss 111 500000 10.0 0.0 20.0
```

The first 3 variables (the `111`) are ICALC, NEWRUN and IPRNT.

	ICALC declares the type of calculation.
ICALC = 0	Continuation of a Monte Carlo simulation.
ICALC = 1	Initial set-up for a Monte Carlo simulation. Origins of initial solute and solvent coordinates are declared in the parameter file (Section 9).
ICALC = 2	Energy minimization for the solute(s) in the gas-phase or single-point QM calculation (Section 15.2).
ICALC = 3	Repetitive energy minimizations are performed for the solute(s) as either one chosen dihedral angle is varied over 360° (dihedral angle driving) or a potential surface is scanned. (Sections 15.3 & 15.4.)
ICALC = 4	Perform normal mode frequency calculations for the solute(s) (Section 15.5).
ICALC = 5	Conformational search by random variation of internal coordinates (Section 15.6).
ICALC = 6	Stochastic conformational search (Section 15.6).
ICALC = 7	Perform single-point force-field calculation.
NEWRUN = 0	Usual case; continue averaging in a Monte Carlo simulation. If ICALC = 5, a conformational search is restarted from an old up file.

NEWRUN = 1 To begin new MC averaging by starting a new "up file". NEWRUN is set to 1 for only the first run after equilibration has been completed. If NEWRUN is accidentally set to 1 during averaging, the averaging is restarted and the previous runs are lost – though the system is probably now very well equilibrated! If ICALC = 5, a new conformational search starts from scratch.

IPRNT = 0 Print solute coordinates and intrasolute distances.

IPRNT = 1 Print solute coordinates, but no distances (normal choice).

IPRNT = 2 Print solute and solvent coordinates.

IPRNT = 3 Suppress printing of coordinates and many parameters.

The next variable is MXCON (the `500000`) which specifies the number of MC configurations for this run. **Keep it constant during averaging**, *i.e.* after NEWRUN = 1. It is often advisable to perform some very short runs at the beginning for a new system just to check that everything has been set up properly. For example, with a `111` start just run 100 configurations or so (MXCON) and check the system graphically by displaying the resultant plt file. Then restart with `011`, if all is okay. Also, the initial energy may be very high in liquid simulations due to some overly short interatomic contacts. These are typically relieved within *ca.* 10K configurations. It may be advisable not to perform volume moves initially (set NVCHG = 999999 in the parameter file), *e.g.*, for the first 300K of equilibration, since undesirable volume expansion may occur to relieve the short contacts. MXCON can have up to 7 digits.

If ICALC = 2, 3 or 5, MXCON has a different meaning, *e.g.*, for ICALC = 2, MXCON is the maximum number of geometry optimization cycles; for ICALC = 3, MXCON is the atom number for the internal coordinate to be varied during the dihedral angle driving or potential surface scanning; and, for ICALC = 5, MXCON is the number of random starting structures to be generated in the conformational searching.

The final three variables (`10.0 0.00 20.0`) are RC0, RC1 and RC2 which define the reference and two perturbed systems for free energy MC calculations. Normally, they range from 0 to 1, though other values may be convenient for potential-of-mean-force computations, as in the d010cmd file. Geometrical and potential function parameters are scaled linearly between initial and final values indicated in the zmat file as RC goes from 0 to 1. RC0 is for the reference solute and RC1 and RC2 are for the two perturbed solutes that are treated by the program. Thus, perturbations can be made in two directions (double-wide sampling). If a free energy perturbation is being performed, the parameter and Z-matrix files can be set up so that for new increments (change of RC0) only the command file requires change and normally the last in file is copied to be the initial in file for the new increment. If a free energy calculation is not being performed, set RC0 = RC1 = RC2. If ICALC = 3 with dihedral driving, RC0 is the dihedral angle increment in degrees. For potential surface scanning, RC0 is the increment in Å or degrees for the variable, RC1 = 1.0,, 2.0 or 3.0 for varying a bond length, bond angle or dihedral angle, and RC2 = the total number of energy minimizations to be performed.

Complete perturbations can be set up by chaining jobs. The command file then needs three sections: equilibration, averaging, and submission of next job. The latter can be done by copying the current *in* file to be the *in* file for the next perturbation window, and then executing the next command file. For example, under *UNIX*:

```
cp d010in d030in
csh -e d030cmd >& log30 &
```

Then, the d030cmd file executes the next command file, etc. The test job dihpmf/fullpmf illustrates this, and it also provides a single command file for computation of an entire pmf from one job submission. Under *Windows*, the above csh command would be replaced by *call d030*.

9 Parameter File Input

If you are just using the x scripts, you can skip this section.

The parameter file begins with the definition of variables for the simulation. Short descriptions are provided on alternate lines and are mostly self-explanatory. Most parameters will be assigned default values, if a value of zero is specified. To obtain a parameter file for a new problem, either (a) execute the interactive par file generator, pargen (just type pargen), or (b) copy an old parameter file and edit it. The par files for the test jobs were all created by parggen. Parggen asks the user a series of questions that allows facile construction of the par file without the user having to be an expert on all of the following details. Some highlights are:

SVMOD1, STREN Designates the primary solvent choice. The current options from stored boxes are water (TIP3P, TIP4P, or TIP5P), methanol, acetonitrile, dimethyl ether, THF, propane, methylene chloride, chloroform, carbon tetrachloride, argon, DMSO, and NONE (for a gas-phase calculation). Additional, "custom" solvents are obtained by designating SVMOD1 as OTHER and providing a Z-matrix for one solvent molecule (Section 10.13). For solventless or non-Monte Carlo simulations, SVMOD1 and SVMOD2 should be set to NONE. To include GB/SA solvation, declare GBSA or GB/SA and the optional ionic strength STREN in mol/l.

The possible entries are: TIP3P, TIP4P, TIP5P, CH3OH, CH3CN, MEOME, THF, C3H8, MECL2, CHCL3, CCL4, ARGON, DMSO, OTHER, NONE, GBSA

QMNAME, CMNAME, QMSCALE, ICH

IF QMNAME = AM1 or PM3, the solute energetics are evaluated with these methods. For PDDG/PM3, QMNAME = PDG; for PDDG/MNDO, QMNAME = MDG. Leave QMNAME blank for a standard fully MM calculation. CMNAME = CM1, CM3 or MULL (for Mulliken) as the choice of solute charge calculation. QMSCALE is the factor for scaling the computed QM charges - normally 1.08 for neutral solutes and 1.0 for charged ones. Then, the net charge (normally 0, -1, or 1) for the reference, 1st perturbed, and 2nd perturbed solutes are given. These are ICH0, ICH1, and ICH2 in the program.

If QMNAME = AM1S, PM3S, PDGS or MDGS and ICALC = 2, a single-point AM1 or PM3 calculation is performed for the solute. If QMNAME = AM1, PM3, PDG or MDG and ICALC = 2, a full AM1 or PM3 optimization is performed for the solute. In both cases, the resultant atomic charges and generic Lennard-Jones parameters are written to the sum file. This is useful if one wants to then perform MM calculations with these non-bonded parameters. See Section 18.1 for additional information.

Solute, Solvent Origin

The next two entries designate the origin of the initial (ICALC = 1) solute and solvent coordinates. For the solute(s), the options are from the Z-matrix file in Z-matrix (ZMAT), Protein Data Bank (PDB), PDBP (PDB with q, σ, ε appended) or MIND format, from the solute coordinates in the in file (IN) (useful if a gas-phase MC run has been performed to get a low-energy solute structure; use the output in file from the gas-phase run as the current in file), and from maximal mapping of the new system onto the one in the in file (ZIN). For the ZIN option, the solvent coordinates are taken from the in file, while the solute coordinates come from the in and Z-matrix file - the coordinates of the first 3 atoms of each solute are taken without change from the in file and used to build the rest of the solute structures from the Z-matrix. For the solvent, the options for the initial coordinates are from the stored boxes (BOXES), from the coordinates in the supplied in file (IN), or NONE. A custom solvent box can be read from the in file by designating IN; this is useful if one has previously equilibrated a pure custom solvent box.

ICAPAT CAPRAD CAPFK

When a solvent cap is desired, ICAPAT must be non-zero and designates the solute atom that defines the center of the cap. CAPRAD gives the cap's radius from ICAPAT, and CAPFK is the harmonic restoring force constant in kcal/mol-Å^2 for solvent that strays beyond CAPRAD. It is suggested that ICAPAT be placed by itself as a separate solute. The program will recognize this and not attempt to move ICAPAT. Otherwise, if ICAPAT is a regular solute atom, the solute motion will be constrained by the change in the cap potential when ICAPAT is moved unless CAPFK = 0. When a cap is desired, the normal choices of solvent origin are BOXES to carve the cap out of a stored box or IN to read the solvent coordinates from the supplied in file - see Section 14 for details.

OPTIMIZER

For solute optimizations (ICALC = 2 or 3), the choice of methods is Simplex (SIMPLX), Powell (POWELL), Fletcher-Powell (FLEPOW), Hooke-Jeeves (HOOKE), BFGS (BFGS), Conjugate Gradient (CG), or Simulated Annealing (SA).

FTOL FTOL specifies the convergence criterion in kcal/mol between cycles with these optimizers.

NEWZMAT If NEWZMAT = NEWZMAT, the final Z-matrix from the optimization or QM single-point calculation is written to the sum file. In addition, if NEWZMAT = FULL, any automatic designations in the input Z-matrix are fully expanded in the output Z-matrix.

NSAEVL The maximum number of potential energy function evaluations in Simulated Annealing. If it is exceeded, the simulated annealing is terminated.

NSATR The number of iterations before temperature reduction in the Simulated Annealing. After (NSATR*20*number of variables) function evaluations, temperature (SATEMP) is scaled by the factor SART.

SATEMP The initial temperature in simulated annealing. It controls the step length over which the algorithm searches for minima. It is very important to choose a high enough temperature which leads to the global minimum. If this is too small, insufficient searching may occur. To determine an appropriate starting temperature, run a trial with SATEMP = 1.0, SART = 1.5, NSAEVL = 1000, and IPRNT = 3. Since the temperature reduction factor SART is greater than 1.0, the temperature increases and the step length rises as well. The temperature and the step length are printed in the output. Select the temperature which produces a large enough step length.

SART The temperature reduction factor in simulated annealing. *I.e.*, the temperature in the next iteration will be SART * current temperature.

ICSOPT Options in conformational searching (Section 15.6).
ICSOPT(1) ... refine geometry, by imposing stricter convergence limits.

CSETOL Energy criterion for a unique conformer in conformational searching.

CSRMS RMS tolerance in superposition of conformers in conformational searching.

CSRTOL Distance criterion for a unique conformer in conformational searching.

CSREUP Upper bound energy relative to the current lowest minimum in conformational searching.

CSAELO Absolute lower bound energy in conformational searching.

CSAEUP Absolute upper bound energy in conformational searching.

NMODE If NMODE is set to 1 and ICALC is 2, 3, or 4, a normal coordinate analysis is performed (Section 15.5).

NSYM The symmetry number in normal mode analysis.

NCONF The number of equivalent conformers in normal mode analysis.

FSCALE Frequency scale factor in normal mode analysis.

FRQCUT Cutoff frequency in normal mode analysis.

NMOL
IBOX
BCUT IBOX controls the choice of desired solvent box from those listed in Section 16, where IBOX is the number of molecules in the stored box. If the initial (ICALC = 1) solvent coordinates are coming from the IN file, IBOX is automatically set to the number of solvent molecules in the IN file. NMOL designates the total number of solvent molecules to be retained for the present simulation. The program automatically discards the (IBOX- NMOL) solvent molecules with the worst interaction energies with the solute(s). To obtain low initial solute-solvent energies, a rule-of-thumb is to discard an equivalent number of non-hydrogen solvent atoms as there are non-hydrogen solute atoms. For example, if there are 40 such solute atoms and the chosen solvent is chloroform, discard 10 chloroform molecules. For solventless or non-Monte Carlo simulations, both SVMOD1 and SVMOD2 should be specified as NONE and IBOX, NMOL and NMOL2 as 0.

The program will automatically create a solvent box when NMOL = 9999 for ICALC = 1. In this case, set IRECNT = 1 and IZLONG = 1 to center the solute. The box will extend BCUT Å beyond the farthest solute atom in the ±x, ±y, and ±z directions. The disadvantage of this procedure over directly using the stored boxes is that the faces of the box are rough-cut. This leads to a high initial energy, so longer equilibration is required including *ca*. 300 000 configurations without volume moves (make NVCHG large). However, systems with up to 3000 solvent molecules can be created. Solvent molecules with any atom initially within 2.5 Å of a solute atom are removed. The desired box is cut from a very large solvent cube, *ca.* 100 Å on a side, that is generated from 27 images of smaller, stored boxes.

SVMOD2
NMOL2 If a binary solvent mixture is desired, the second solvent choice and number of molecules are declared by SVMOD2 and NMOL2. The binary solvent boxes are created by first generating a box with NMOL SVMOD1 molecules. NMOL2 of these are replaced randomly by SVMOD2 molecules. Thus, the total number of solvent molecules is NMOL consisting of (NMOL-NMOL2) of type SVMOD1 and NMOL2 of type SVMOD2. The initial total energy for the binary systems will be high, so equilibration for *ca.* 500K configurations without volume moves (set NVCHG = 999999) is recommended. Then, set NVCHG = 0 to begin the NPT calculations and equilibrate the volume and energy. **Note**: Energy distributions are for the mixture; however, the radial distribution functions are only for atoms in the first solvent. Also, either SVMOD can be OTHER, but not both. For solventless or non-Monte Carlo simulations, SVMOD1 and SVMOD2 should both be NONE.

NCENT1
NCENT2
NCENTS

The atoms used to define the centers of solutes 1 (NCENT1) and 2 (NCENT2) for the purposes of the cutoff procedures designated by ICUT below. If there is only one solute, and ICUT=0, NCENT2 can be 0 (ignored) or non-zero, which causes the average of the positions of NCENT1 and NCENT2 to be used as the reference point. *E.g.*, for butane (C1-C2-C3-C4), if NCENT1 = 2 and NCENT2 = 3, the middle of the C2-C3 bond will be the reference point. The NCENT for solutes above 2 is taken as the first atom of the solute. NCENTS is the atom used in a custom solvent molecule for cutoffs; otherwise, atom 1 is used.

ICUTAS

For custom solvents, the solvent atoms that are used for the solvent-solvent and solute-solvent cutoffs. The maximum number is five.

NROTA1
NROTA2

The atoms that solutes 1 and 2 are rotated about for the total (rigid-body) rotations. Often, these are set equal to NCENT1 and NCENT2. For free energy perturbations in which geometries are varied, NROTA1 and NROTA2 should be chosen, if possible, to be atoms whose positions are identical in all three solutes images (the reference and two perturbed ones). Otherwise, the images' relative orientations will change during the simulations. The NROTA for solutes above 2 is taken as the first atom of the solute.

ICUT Designates the cutoff procedure for the solute-solvent interactions based on distances between solute atoms and atom NCENTS of the solvent molecule. The options are described in function WXPOT. Normally,

ICUT = 0 For small, neutral, single solutes. Bases cutoff on distance to the average of atoms NCENT1 and NCENT2 in the solute. If the distance is less than SCUT, the entire solute–solvent molecule interaction is included.

ICUT = 1 Not a valid choice.

ICUT = 2 A possibility for large solutes. Bases cutoff on all solute atoms except type -1 dummy atoms. If any solute–solvent atom 1 distance is less than SCUT, the interaction between the entire solute and solvent molecule is included. Slower to evaluate than ICUT = 4 or 5.

ICUT = 3 Bases cutoff on shortest distance to NCENT1 and NCENT2. Much faster than ICUT = 2, typically. If the NCENT1 or NCENT2 to solvent atom 1 distance is less than SCUT, the interaction is included. Often the choice for two small solutes.

ICUT = 4 Similar to ICUT = 3 except all solute atoms designated in the ICUTAT array (maximum of 15) are used. This or ICUT = 5 are the preferred procedures for large solutes and represent a compromise between ICUT = 2 and ICUT = 3.

ICUT = 5 Similar to ICUT = 2 except only non-hydrogen atoms of the solute are used for the cutoff. Often the BEST CHOICE.

Note: For ICUT = 2 - 5 if there are multiple solutes, they are treated independently. Thus, if the solvent molecule is within SCUT for one solute, but not another, only the interaction with the nearer solute is included. If there are more than two solutes, use ICUT = 2, 4, or 5. With ICUT = 4, at least one atom in the ICUTAT array should be from each solute.

ICUTAT Solute atoms used for ICUT = 4 (maximum 15 atoms).

IRECNT To recenter the solute(s) in the middle of the simulation box, set IRECNT = 1. All molecules are translated so the energy is unaffected. This is just cosmetic.

INDSOL Set INDSOL = 1 to let the solutes move independently. If INDSOL = 0, the first two solutes only will be translated in tandem, which is useful for some pmf calculations.

IZLONG Set IZLONG = 1 to have the solute(s) oriented with their longest axis along the Z-direction. This is only performed at the beginning of a simulation, *i.e.*, when ICALC = 1. It has no effect otherwise.

MAXOVL Set MAXOVL = 1 for the free energy perturbations to have the perturbed solutes maximally overlaid on the reference solute. This uses the QTRF subroutine to perform least squares fits after every solute move. It can help reduce noise in the free energy results and takes little CPU time.

NOXX
NOSS For use of multiple solute or solvent copies. If NOXX = 1, solute-solute interactions are not evaluated except with solute 1. If NOSS = 1, solvent-solvent interactions are not evaluated; only solute-solvent and internal energies are evaluated.

NOBNDV
NOANGV For some equilibrations - if NOBNDV = 1, variable bonds are not varied; if NOANGV = 1, variable angles are not varied.

NONEBN If NONEBN = 1, only bonds declared as variable and additional in the Z-matrix are considered to be covalent bonds for determining covalent neighbors. This is the normal case. Otherwise, bonds are determined by interatomic distance, which may be misleading for a high-energy structure.

NRDF
NRDFS
NRDFA These are used to designate the requested solvent-solvent and solute-solvent radial distribution functions. See sample par files. The atoms designated in NRDFA need to be in ascending order, e.g.,

```
 2  2  2  2  5  5  7  7  9
 1  2  3  5  1  2  1  2  1
```

NVCHG
NSCHG An attempt to change the system's volume is made every NVCHG configurations. If an NVT (constant volume) simulation is desired, set NVCHG = 999999. A solute move is attempted every NSCHG configurations. NVCHG and NSCHG will be determined automatically, if they are input as 0. For solutes with many internal variables, smaller values of NSCHG are appropriate.

MAXVAR Defines the maximum number of variable bond lengths, bond angles, and dihedral angles that will be varied on an attempted solute move (MAXVAR of each type). The default is 15 if MAXVAR = 0.

NCONSV The coordinates for the full system are written to the "save' file every NCONSV configurations during a MC simulation in the same format as for an "in" file.

USEFUL OPTION: If NCONSV = -1, the Z-matrix file for the lowest-energy structure from the MC run is written to the "save" file. The latter feature is useful for doing a high-T MC search to find the lowest-energy structure.

NBUSE Set NBUSE = 1 to use neighbor lists for evaluating the solvent-solvent interactions. This saves computer time for large systems. If the neighbor lists overflow, a message is sent and NBUSE will need to be set to 0.

NOCUT Set NOCUT = 1 to *not* add the cutoff correction for the solvent-solvent Lennard-Jones interactions for solvents other than TIPnP water.

NOSMTH Set NOSMTH = 1 to *not* smooth the solute-solvent and solvent-solvent interactions near the cutoff.

WKC Constant for the preferential sampling using $1/(r^2 + WKC)$ weighting. *WKC* needs to increase as the number of molecules increases, otherwise there will be gradual volume expansion in NPT simulations. Some typical values of *WKC* are 150 for 216 or 250 water molecules, and 250 for 267 chloroform molecules.

RDELS1
RDELS2
ADELS1
ADELS2 These are the ranges for attempted translations in Å and rigid body rotations in degrees for the first and second solutes. The ranges for additional solutes are automatically determined and dynamically adjusted to give *ca.* 50 % acceptance rates for attempted moves. If the ranges are 0, the solutes will not be translated and/or rotated.

LHTSOL
TLHT Solute number LHTSOL will be heated at temperature TLHT. The sampling for the other solutes and solvent are at temperature T.

RCUT The cutoff distance for solvent-solvent interactions based on the NCENTS-NCENTS (usually atom 1 – atom 1) distances. All interactions are quadratically feathered to 0 between RCUT and RCUT - 0.5 Å.

SCUT The cutoff distance for solute-solvent interactions based on the NCENT to NCENTS distances depending on ICUT. Feathered as above.

CUTNB The cutoff distance for intrasolute non-bonded pairs. Atom pairs separated by more than this distance are not included in the non-bonded pairs list. The list is recomputed when a new job is executed.

EWALD Set EWALD = 1 for use of Ewald summation (Section 17). This is only proper for periodic boundary conditions and OPLS energy evaluation.

T, P The temperature in °C and the external pressure in atm.

TORCUT Individual torsional energy terms are truncated at TORCUT kcal/mol. This may be useful for equilibrating a flexible system. If TORCUT = 0.0, no cutoff is applied

DIELEC The dielectric constant for solute optimizations and continuum simulations (Section 15). DIELEC is set to 1 for all simulations with explicit solvent molecules.

DIELRF Interactions beyond the cutoffs with explicit solvent can be treated with the reaction field procedure by specifying a dielectric constant for the reaction field > 1. See *J. Phys. Chem.* **1995**, *99*, 17956; eq (1) is implemented. This is only appropriate for solutions of neutral molecules. The usual feathering of the intermolecular interactions within 0.5 Å of the cutoffs is not done.

SCL14C
SCL14L The Coulombic and Lennard-Jones scaling factors for 1,4-intramolecular non-bonded interactions. If both are > 99999, then 1,4 interactions are ignored. The Coulombic and Lennard-Jones 1,4-interaction energies are divided by SCL14C and SCL14L, respectively. The default value is 2.0 for both, which is proper for the OPLS-AA force field. ***Important***: For the united-atom force field in the oplsua.par file, use SCL14C = SCL14L = 99999.99

PLTFMT The format for the solute and solvent coordinates written to the PLT file. The options are PDB, PDB2, PDBB, MDLMO, MDLSD, and MIND. PDB generates a Protein Data Bank format file with dummy atoms removed that is suitable for displaying with programs like ChemEdit, RasMol, Sybyl, MacroModel or Midas. PDB2 generates a PDB file including the coordinates for the reference solute and the first perturbed solute. This is useful for visualizing the structural change for the perturbation. PDBB provides a PDB file with the BOSS atom types appended after the solute coordinates. Such a file can be read by BOSS to begin a simulation (Section 11). MDLMO or MOL yields a file in MDL Molfile format, MDLSD yields a file in MDL SD format, and MIND yields a file formatted for display with the MindTool program. Again, dummy atoms are removed.

ISOLEC The solute-solvent and solute-solute energies are decomposed into their Coulomb and Lennard-Jones components for solute ISOLEC. The default is solute 1. These components are useful for linear response calculations, see: H. A. Carlson and W. L. Jorgensen, *J. Phys. Chem.* **99**, 10667 (1995).

The parameter file then lists non-bonded potential function parameters (q, σ, ε) by atom type. Many of the command files automatically append this information from the oplsaa.par file to the end of the top part of the par file. If the user provides a complete par file, new atom types can be defined at the end of the current list. *Only parameters that are used for the present calculation need to be listed. The type numbers do not have to be sequential.* However, since parameters for the stored solvents occur up to atom type 126, the listing of parameters is often not truncated before this point.

The final entries in the parameter file are the Fourier coefficients for different dihedral angle types. Again, new types can be added at the end of the list and only the parameters for the current calculation need to be listed. The torsional potential for angle ϕ is defined as

$$V(\phi) = V1(1+\cos\phi)/2 + V2(1-\cos 2\phi)/2 + V3(1+\cos 3\phi)/2 + V4(1-\cos 4\phi)/2$$

V4 is almost always zero. The input format for specifying the torsional parameters for an angle is:

```
Type   V1      V2      V3      V4       a1-a2-a3-a4

I3  2x F8.4 2x F8.4 2x F8.4 2x F8.4 4x  A2-A2-A2-A2

as in

123xx12345678xx12345678xx12345678xx12345678xxxx
```

```
006  1.711     -0.500     0.663     0.0         CT-CT-CT-O?
```

This example is for the CT-CT-CT-OS or CT-CT-CT-OH torsion in alkyl ethers or alcohols. The AMBER atom type quartet at the end of the line is needed for proper automated assignment of dihedral types. The quartet can be entered either as w-x-y-z or z-y-x-w. The ? is a wildcard that is used in the matching algorithm; the priority is exact match > ? > mismatch. Non-bonded (Coulomb plus Lennard-Jones) interactions are also included within a solute for >1,3 interactions, when there is internal motion.

The complete lists of OPLS united-atom and all-atom non-bonded and torsional parameters are in the files *oplsua.par* and *oplsaa.par*, respectively. These files can be appended to the top section of the parameter file, which is output by pargen, to yield a complete parameter file. The files containing bond stretching and angle bending parameters are *oplsua.sb* and *oplsaa.sb*.

10 Z-matrix File Input

The Z-matrix file contains the specifications for the initial setup of the solute geometries. It also gives the atom type designations and the 3-character labels for the atoms that are used in printing. For gas-phase energy minimizations, the variables to be optimized are declared. Any geometry changes for an MC-FEP calculation are also designated as well as any internal variables (bond lengths, bond angles, dihedral angles) to be varied. And, finally, domain definitions are made for evaluation of intrasolute >1,3 – nonbonded interactions when internal variables are sampled over. It is important to realize that the Z-matrix is typically used to build the coordinates of the solute molecules when an MC solute move is attempted. A solute move for a solute with internal variations consists of (1) randomly selecting which of the solutes is going to be moved (obviously this step can be skipped if there is only one solute), (2) translating the first three atoms of the solute randomly in all three Cartesian directions, (3) rotating the first three atoms of the solute randomly about one randomly chosen Cartesian axis, and (4) then constructing the rest of the solute from the first three atoms by applying the specifications in the Z-matrix for the remaining atoms of only that solute. Changes in the internal variables are incorporated when the solute is rebuilt from the Z-matrix. If the solute does not have internal variations, steps 2 and 3 are applied to all atoms and step 4 is skipped. For MC perturbation calculations, Z-matrices are maintained internally for the two perturbed solute systems (each of which may have multiple fragments). The moves of these perturbed solutes are performed simultaneously with the move of the reference solute system. The same displacements are employed so maximum overlap of the reference and perturbed solutes is maintained.

Variations of this procedure are possible by appropriate designations in the parameter file. A handy one is to force the first two solutes to be translated in tandem by setting INDSOL = 0. This is necessary if a potential of mean force is being determined as a function of an intersolute distance. In this case, the two atoms that define the distance need to be declared as NROTA1 and NROTA2. The solutes will then be rotated about these two atoms which will maintain their specified separation that is also maintained by the tandem translation. The maximum number of atoms in the Z-matrix is 1000. The maximum number of variable and additional dihedral angles is 1200, bond lengths 950 and bond angles 950.

When constructing Z-matrices, work from the middle of the molecule outward. This will result in smaller structural changes for variations in dihedral angles and lead to better sampling. I.e., place the first atoms including dummies near the middle of the molecule. Avoid starting the Z-matrix at one end of the molecule.

The program checks solute Z-matrices for common errors and inconsistencies. If potential problems are found, they are noted in the ot file and should be investigated.

The aazmat and peptide directories contain hundreds of all-atom Z-matrices for molecules, complexes, and all common amino acids as dipeptides. Z-matrices for new projects are usually obtained by building from one of these, by using the molecule builder in ChemEdit, or by conversion of a PDB or mol file by the xPDBZ or xMOLZ scripts.

The following describes the format for the input:

```
Line 1: TITLE USED IN PRINTING (A30)
```

Subsequent lines:

```
ATOM #I, SYMBOL, TYPE, FINAL TYPE, J, RIJ, K, TH(IJK), L, PHI
I4  1X   A3 1X  I4   1X  I4 1X  3(I4,F12.6)

LAST LINE OF Z-MATRIX: BLANK LINE
```

Additional input read by routines ZSTART and NBPAIR is in 10 groupings separated by lines with the first 3 columns blank:

1. Geometry Variations – specify 1 per line (for perturbations):

```
ATOM NUMBER, PARAMETER (1,2, OR 3), FINAL VALUE
     I4              I4              F12.6
```

Then: `BLANK LINE`

2. Bonds to be varied - specify the atom number from the Z-matrix, one per line (`I4`) or a contiguous range of atoms (`I4-I4`), *e.g.*, 0003-0008 (the hyphen is needed):

```
ATOM NUMBER  - ATOM NUMBER
     I4      -      I4
```

Then: `BLANK LINE`
or to make a bond length equivalent to another, use an = in column 5, e.g.,

```
0010=0005
```

Place any such equivalence entries at the end of the list of variable bonds.

3. Additional Bonds to include in the energy evaluation, one per line:
(Note these entries are not needed in a QM/MM calculation.)

```
ATOM 1, ATOM 2
  I4      I4
```

Then: `BLANK LINE`

4. Harmonic Restraints for solute atom pairs, $E = K(R-R0)^2$, one per line (maximum of 90):

```
ATOM I, ATOM J, R0, K0, K1, K2
  I4      I4        4F10.4

or for restraining atom I to a point x, y, z, one per line:
ATOM I, 9999, K, x, y, z
  I4      I4     4F10.4

or for a flat-bottomed harmonic used for NOE restraints, one per line:
ATOM I, ATOM J, -1.0, R-low, R-high, K
I4      I4        4F10.4
```

Then: `BLANK LINE`

5. Bond Angles to be varied - specify the atom number from the Z-matrix, one per line or a contiguous range of atoms (I4 – I4):

```
ATOM NUMBER  -  ATOM NUMBER
     I4      -       I4
```

Then: BLANK LINE

or to make a bond angle equivalent to another, use an = in column 5, e.g.,

```
0010=0005
```

Place any such equivalence entries at the end of the list of variable angles.

6. Additional Bond Angles to include in the energy evaluation, one per line:
(Note these entries are not needed in a QM/MM calculation.)

```
ATOM 1, ATOM 2, ATOM 3
  I4      I4      I4
```

Then: BLANK LINE

Or for automatic determination by the program, just declare AUTO in columns 1-4:

AUTO

Then: BLANK LINE

7. Specify Dihedral Angles to be varied, one per line:

```
ATOM NUMBER, INITIAL TYPE, FINAL TYPE, RANGE
     I4            I4            I4      F12.6
```

Then: BLANK LINE

Or, for automatic type and range assignments or for a QM/MM calculation, one can just list the dihedral angles to vary either singly and/or as a contiguous range, e.g., 0005-0014:

```
ATOM NUMBER  -  ATOM NUMBER
     I4      -       I4
```

Then: BLANK LINE

or to make a dihedral angle equivalent to another, use an = in column 5, e.g.,

```
0010=0005
```

Place any such equivalence entries at the end of the list of variable dihedrals.

8. Specify Additional Dihedral Angles to be included in the evaluation of the intramolecular energy, one per line: (Note these entries are not needed in a QM/MM calculation.)

```
ATOM 1, ATOM 2, ATOM 3, ATOM 4, INITIAL TYPE, FINAL TYPE
  I4     I4      I4       I4         I4            I4
or
AUTO
```

Then: BLANK LINE

Note: for additional bond and dihedral angles, one can mix AUTO with specific designations.

To harmonically restrain dihedral angles, make additional entries:

```
ATOM 1, ATOM 2, ATOM 3, ATOM 4, 500, PHI0   (6I4)
```

Type 500 in the par file is reserved for his purpose. The force constant in kcal/mol-rad^2 is the V0 entry - see oplsaa.par. $E(PHI) = K(PHI-PHI0)^2$.

9. Domain Definitions for exclusion in non-bonded pairs calculations - only if internal variables are sampled. This reduces the number of non-bonded pairs that need to be computed for the solutes. Enter Domain 1 and Domain 2 pairs, one set per line - 4 atoms are needed:
(Note these entries are irrelevant in a QM/MM calculation.)

```
1ST ATOM DOM1, LAST ATOM DOM1, 1ST ATOM DOM2, LAST ATOM DOM2 (4I4)
```
Then: BLANK LINE

10. Bond lengths, bond angles and dihedral angles to be sampled for Monte Carlo conformational searching, one per line (this is optional; the program will select key torsions, if not specified):

```
ATOM NUMBER, PARAMETER (1,2 OR 3), MINIMUM, MAXIMUM
     I4          I4                   F12.6     F12.6
```

or for stochastic conformational searching:

```
NFIX, RAD, PUSH, HWIDTH  (36X, I4, 3F10.4) – see section 15.6.
```
Then: Final BLANK LINE

11. After the "Final BLANK LINE", there can be optional designation of non-bonded parameters for atoms whose charges have normally come from a QM single-point calculation, e.g., using xAM1SP, xPM3SP, or xPDGSP for neutral systems, xPDGSP+ for a cation, etc. For example, in the *aazmat* directory, if one executes *xAM1SP meoh* , the output *sum* file is the following:

```
        Methanol                                          Tot. E =    -55.0866
   1 O    800  800    0    0.000000   0    0.000000   0    0.000000         0
   2 DUM   -1    0    1    0.500000   0    0.000000   0    0.000000         0
   3 DUM   -1    0    2    0.500000   1   90.000000   0    0.000000         0
   4 H    801  801    1    0.945698   2   90.000000   3  180.000000         0
   5 C    802  802    1    1.411870   4  108.989752   2  180.000000         0
   6 HC   803  803    5    1.090450   1  110.239001   4  179.999996         0
   7 HC   804  804    5    1.090404   1  110.549523   6  119.850735         0
   8 HC   805  805    5    1.090404   1  110.549526   6  240.149263         0
                    Geometry Variations follow
                    Variable Bonds follow         (I4 or I4-I4)
0004-0008
                    Additional Bonds follow       (2I4)
                    Harmonic Constraints follow
                    Variable Bond Angles follow   (I4 or I4-I4)
0005-0008
                    Additional Bond Angles follow (3I4)
AUTO
                    Variable Dihedrals follow     (3I4,F12.6)
0006-0008
                    Additional Dihedrals follow   (6I4)
AUTO
                    Domain Definitions follow     (4I4)
                    Conformational Search         (2I4, 2F12.6)
                    Final blank line

 Final Non-Bonded Parameters for QM (AM1 CM1Ax1.14) Atoms:

 800  8 OH  -0.586957  3.120000  0.170000
 801  1 HO   0.406869  0.000000  0.000000
 802  6 CT  -0.043638  3.500000  0.066000
 803  1 HC   0.074575  2.500000  0.030000
 804  1 HC   0.074575  2.500000  0.030000
 805  1 HC   0.074575  2.500000  0.030000
```

Note that the usual OPLS-AA atom types have been replaced by types 800-805, whose non-bonded parameters including the 1.14*CM1A charges have been designated at the end of the new Z-matrix. Atom type numbers 800-899 and 9500-9999 have been reserved for this purpose, and the corresponding non-bonded parameters are read from the Z-matrix file rather that the *oplsaa.par* file.

Some explanatory details:

10.1 Atom Input

Note: the first atom is placed at the origin (0,0,0), the second atom is on the +x axis and the third atom is in the xy plane. All of the rectangular solvent boxes have the long axis as z. TYPE and FINAL TYPE specify the initial (RC0 = 0) and final (RC0 = 1) atom types for atom i. The corresponding non-bonded potential function parameters are retrieved from the par file. If TYPE = FINAL TYPE, FINAL TYPE can be left blank or shown the same as TYPE. For systems with **multiple solutes**, separate them with a line having "TERZ" in columns 1-4. Also note there are two types of dummy atoms; one with type -1 is a true dummy that is only used in constructing the solute coordinates from the Z-matrix, while a dummy with type 100 has $q = \sigma = \varepsilon = 0$, but can be converted in an FEP calculation to a real atom with $q \neq \sigma \neq \varepsilon \neq 0$.

10.2 Geometry Variations

This designates how geometrical parameters are to be changed for a perturbation. The ATOM NUMBER corresponds to the numbering in the Z-matrix input including dummy atoms. PARAMETER refers to bond length (1), bond angle (2), or dihedral angle (3). The initial parameter value is in the Z-matrix and refers to RC (or λ) = 0. The final parameter value is specified here and refers to RC = 1. The parameter value is scaled linearly for intermediate values of RC:

$$\theta(RC) = RC \times (\theta_{final} - \theta_{initial}) + \theta_{initial}$$

Note: in the command file, the RC values can be outside the (0,1) range. For example, in computing a potential of mean force as a function of a distance, one could state in the Z-matrix file that the initial distance is 0 Å for RC = 0 and 1 Å for RC = 1. Then, if RC = 4, for example, in the command file, the program knows to set the system up with the distance as 4(1-0) + 0 = 4 Å. However, for molecular mutations, *e.g.*, $CH_3OH \rightarrow CH_3CH_3$, one just lets the mutation run from RC = 0 to 1 since the potential function parameters need to start as methanol and end up as ethane as RC goes from 0 to 1.

10.3 Bond Length Variations

Bond lengths in the Z-matrix can be varied during the Monte Carlo sampling by specifying the atom that is attached by the bond. Specify one atom per line in `I4` format or a range of atoms, I-J in format `I4-I4` (include the `-`); all bonds attaching atoms I through J are included. An = in column 5 specifies a forced equivalence, e.g., 0034=0008, the bond defining atom 34 is the same length as the one for atom 8. Only 0008 and not 0034 should be designated as variable. The harmonic bond stretching parameters are retrieved automatically from the file oplsua.sb or oplsaa.sb. If the atom types for the bond are changing during a free energy calculation, the parameters for the bond in the perturbed solutes are also determined automatically. The ranges for the bond length variations are determined by the program. Only a random subset of variable bonds is varied on each attempted solute move (see subroutine MOVMOL for details). Any missing parameters are flagged in the output file; default values are assigned, but proper additions should be made to the oplsua.sb or oplsaa.sb file.

10.4 Additional Bonds

These are bonds that are not explicit in the Z-matrix, but that are to be included in the energy evaluation. The typical case is a ring closure bond. Specify the two atoms in the bond

(2I4 format), one pair per line. Again, the parameters are retrieved automatically. No entries are needed for QM/MM calculations.

10.5 Harmonic Restraints

Bond Restraint. Harmonic restraints can be included between any pairs of solute atoms (maximum of 90 pairs). For the I-J pair, R0 specifies the reference separation in Å, and K0, K1 and K2 are the force constants in kcal/mol-Å^2 for this restraint in the reference solute and the two perturbed solutes, if any. Each restraint causes the following energy term to be added to the total energy where K = K0, K1 and K2 for the three solutes:

$$EBC = K(RIJ - R0)^2$$

EBC is added to the solute-solute energy, EXX. Even if there is only one solute, the restraints energy is included in the total energy.

The restraints are useful in some special contexts. For example, in mutating one host/guest complex A···B to another C···D, one could perturb from A···B to A—B to C—D to C···D where ··· means the two molecules are completely free, while the introduction of the harmonic constraints is symbolized by —. This keeps the solutes from floating apart during the intermediate stages of the mutation. Other uses, such as restrained optimizations are possible.

Position Restraint. Atoms can also be restrained to a fixed point in Cartesian space. This can be useful for FEP calculations in which a solute is being annihilated - see J. Hermans & L. Wang, *J. Am. Chem. Soc.* **119**, 2707 (1997) and N. McDonald et al., *J. Am. Chem. Soc.* **120**, 5104 (1998). In this case the format for the entry in the Z-matrix file specifies the chosen atom (I), atom J is 9999, and then the force constant and the x, y, and z coordinates for the fixed point are specified. The contribution from each such restraint is then given by EBC = KR^2, where R is the distance between atom I and the fixed point. IRECNT is set to 0, independent of its value in the parameter file, so that the solutes are not automatically recentered at the end of each run, as this would shift atom I relative to the fixed point.

Alternatively, atom J can be restrained to the average of the positions of 1, 2, or 3 other atoms. In this case, specify atom I = 9999, atom J is the atom to be restrained, R0 is the force constant, and K0, K1, K2 are the numbers for the three atoms (these are specified as floating point, e.g., 23.0 for atom 23, and converted to integers by the program).

NOE Restraint. A flat-bottomed harmonic can also be applied between atoms I and J:

$$EBC = K(RIJ - Rhigh)^2 \quad \text{for } RIJ > Rhigh$$

$$EBC = 0 \quad \text{for } Rhigh > RIJ > Rlow$$

$$EBC = K(RIJ - Rlow)^2 \quad \text{for } RIJ < Rlow$$

10.6 Bond Angle Variations

Bond angles in the Z-matrix are varied by specifying the atom for which the angle appears in the Z-matrix or a range of bond angles I-J (I4-I4). The parameters for the harmonic bend are retrieved and the range for attempted variations of the angle is computed automatically. Only a random subset of the variable angles is varied on each attempted solute move. Any missing parameters are flagged in the output file.

10.7 Additional Bond Angles

As for bonds, additional bond angles can be included in the energy evaluation. Specify the three atoms in order x-y-z where y is the central atom, one triple per line. No entries are needed for QM/MM calculations. **Note:***** The first two bonds and intervening angle for each solute can not be variable. Start with two -1 dummies, if this is a problem.*******

The program will automatically determine the additional bond angles by designating AUTO in columns 1-4. In this case, central atoms in explicitly declared variable and additional bond angles and atoms in additional bonds are processed such that all bond angles about these atoms will be included in the MM energy evaluation. The angles that have been added are listed in the ot files.

10.8 Dihedral Angle Variations

For dihedral angles that are sampled, specify the atom set up by the dihedral angle in the Z-matrix and the dihedral angle INITIAL TYPE number whose corresponding For MM calculations, Fourier coefficients must be available in the parameter file. INITIAL TYPE is the initial value for RC = 0 and FINAL TYPE is the type for RC = 1. Often, INITIAL TYPE = FINAL TYPE and FINAL TYPE can be set equal to INITIAL TYPE or 0. However, if an atom in the dihedral angle changes type, the dihedral type may change too. If both INITIAL and FINAL TYPE are 0, the torsional potential is taken as 0, though the angle is still varied.

Also for MC calculations specify the RANGE (PDEL) for variation of the dihedral angle in degrees. Attempted variations of the dihedral angle will be made within ± RANGE degrees of the current value of the dihedral angle.

Alternatively, the program will automatically assign the dihedral types and ranges by designating a group of dihedrals to vary, e.g., 0004-0025, or for a single angle designate, e.g., 0004-0004. The - is needed as the key that auto assignment is desired. Specifically assigned dihedrals and auto ones can be mixed. The assignments are made from the corresponding quartet of AMBER atom types. If there are missing parameters or multiple matching quartets, they are noted in the ot file along with the program's processing decision. Similarly, for QM/MM calculations, the atom numbers for the variable dihedral angles just need to be given as single entries (e.g., 0005) or as a range (e.g., 0005-0014). The TYPE entries are irrelevant, while if a range PDEL for a single entry is specifed, it is used. The default ranges are ±2 to ±10 degrees depending on the V2 value and whether the angle is or is not in a ring (see subroutine PARAM, if interested).

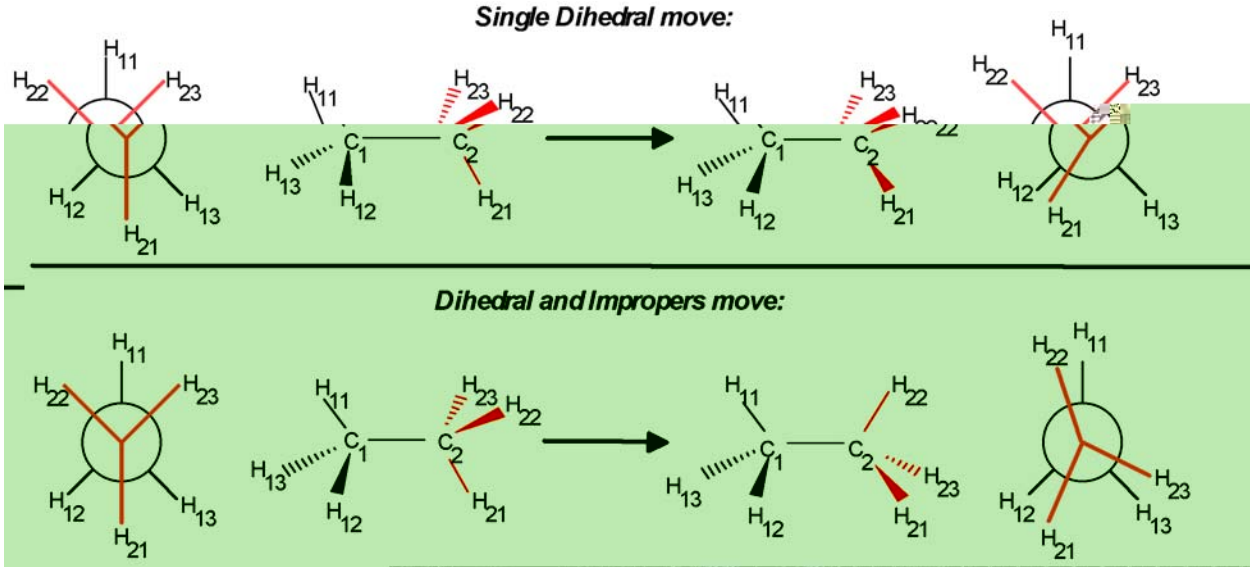
Options for a single angle:

0004-0004	program determines type and range for angle 4
0004 -1 -1 10.0	program determines type and uses ±10° for the range
000400230023 10.0	full specification by user for angle 4 as type 23 with range ±10°

Improper dihedrals may be defined in the atom input section of the Z-matrix to ease sampling of rotation around single bonds. The utility of this technique can be seen in the rotation around the $–C_1–C_2–$ bond in ethane as shown in the scheme below. The top portion illustrates a typical motion done by a change in the dihedral angle $H_{21}–C_2–C_1–H_{11}$ when the remaining hydrogens are defined by normal dihedrals ($H_{22}–C_2–C_1–H_{11}$, $H_{23}–C_2–C_1–H_{11}$, etc.) As shown, the motion of H_{21} changes its bond angles to H_{22} and H_{23} rather dramatically, with a large increase in the energy and almost surely the rejection of the motion by the Metropolis algorithm. A smaller range of motion will be needed which will decrease the efficiency of sampling. The bottom part of the scheme shows the effect of the same change in the $H_{21}–C_2–C_1–H_{11}$ dihedral

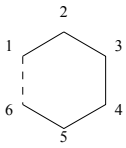
when both H_{22} and H_{23} are defined using improper angles (H_{22}–C_2–C_1–H_{21}, H_{23}–C_2–C_1–H_{21}, etc.). All the bond angles around C_2 are kept constant, and the entire methyl group rotates. This motion is energetically more economical and has a better probability of being accepted. To complete the sampling for H_{22} and H_{23}, small variations (ca. 3°) in the dihedrals H_{22}–C_2–C_1–H_{21} and H_{23}–C_2–C_1–H_{21} can be specified. Many examples can be found in the aazmat directory, e.g., see ethane.z.

***Very important ***: In cases where sampling includes internal variations, the internal changes are made assuming the first 3 atoms of each solute can be used to build the rest of the solute from the Z-matrix. Do not use atoms from any solute to define other than the first three atoms of another solute and the internal coordinates for the first three atoms of any solute can not be varied except for solute optimizations (ICALC = 2). We often use 2-3 dummy atoms at the beginning of a Z-matrix. Autozmat is a convenient way to generate good Z-matrices from coordinate files.



10.9 Additional Dihedral Angles

These are dihedral angles that are to be included in the evaluation of the intramolecular energy, but that are not defined (or varied) explicitly in the Z-matrix. Such angles occur in flexible rings. For the carbon framework of cyclohexane, only 3 dihedral angles are necessary to define the atomic positions in the Z-matrix, but all 6 should be included in the energy evaluation. The 3 additionals can be listed or use AUTO. No entries are needed for QM/MM calculations.



Dihedrals explicit in the body of a Z-matrix which would be declared variable and assigned appropriate potentials:

```
4-3-2-1, 5-4-3-2, 6-5-4-3
```

Additional dihedrals to be included in the energy evaluation:

```
1-6-5-4, 2-1-6-5, 3-2-1-6
```

A dihedral angle phi can also be harmonically restrained by declaring it as an additional dihedral, specifying the initial type as 500 and the final type as the reference value for phi, phi0,

as an integer in degrees (0-360). This is useful for NMR structure refinements. V4 for entry 500 gives the force constant in kcal/mol-rad^2. E(phi) = K(phi-phi0)^2.

10.10 Domain Definitions

For MM calculations, the domain definitions cause intrasolute non-bonded interactions **not** to be evaluated between any atoms in domains 1 and 2. This was typically done when parts of a solute, e.g., an aromatic ring, are treated as rigid to avoid the pointless calculation of associated non-bonded interactions. It is normally not done now using the OPLS-AA force field and fully flexible molecules. Note domain 1 can equal domain 2. If there are internal motions, all other **intra**solute non-bonded interactions will be included and listed in the ot file. Also, some united atom torsion parameters include the non-bonded interactions. For example, with united-atom *t*-butyl alcohol if the torsion for the H on O is included (type 54, V3 = 0.65 kcal/mol) defined as C-C-O-H, this Fourier term fully specifies the torsional potential. However, non-bonded interactions with the H would be added. These can be excluded by the appropriate domain designations or by setting SCL14C = SCL14L = 99999.99 in the par file. Check the non-bonded pairs list in the ot file to verify that the proper exclusions are made.

10.11 Conformational Searching

For Monte Carlo conformational searching, the variable bonds, angles, and dihedral angles that are sampled are declared. If there are no entries, the program will determine suitable torsion angles to be varied that are not in rings. Conformers of a single molecule or alternative geometries for a complex are usually sought. The choices here must also appear in the corresponding variable bonds, angles, or dihedral angles section of the Z-matrix. Otherwise, they are ignored. The ATOM NUMBER and PARAMETER are the same as described in Geometry Variations above. The MINIMUM and MAXIMUM values are the lower and upper bounds of the variables that are allowed in generating starting structures for finding conformers. For a dihedral angle, the bounds are normally 0° and 360°; if listed as 0° and 0°, defaults of 0° and 360° are used. *For details on conformational searching, please see section 15.6.*

10.12 Sample Z-matrix

The following is a simple example of a complete Z-matrix file for an MC-FEP mutation with only one solute using the united-atom force field.

```
Column numbers:
12345678901234567890123456789012345678901234567890123456789012345678901234567890
--------------------------------------------------------------------------------
MeSH to MeCN mutation           Title (A30)
0001 S/N 0083 0094
0002 M/C 0088 0095 0001 1.82
0003 H/X 0087 0100 0001 1.34      0002  96.0
0004 X/M 0100 0096 0002 0.50      0001 180.0     0003 0.0
             Geometry Variations follow         (2I4, F12.6)
00020001 1.157
00030001 0.50
00040001 1.458
             Variable Bonds follow              (I4 or I4-I4)
             Additional Bonds follow            (2I4)
             Harmonic Restraints follow         (2I4, 4F10.4)
             Variable Bond Angles follow        (I4 or I4-I4)
             Additional Bond Angles follow      (2I4)
             Variable Dihedral Angles follow    (3I4, F12.6 or I4-I4)
             Additional Dihedral Angles follow  (6I4)
             Domain Definitions follow          (4I4)
             Conformational Searching           (2I4, 2F12.6)
             Final blank line.
--------------------------------------------------------------------------------
```

In this example, the S becomes the N, the CH_3 group becomes the C, the H is retracted and vanishes, and the CH_3 group of CH_3CN is grown out from the C. The geometry in the Z-matrix corresponds to RC = 0 (CH_3SH), while the final atom types (94, 95, 100, 96) and bond lengths (1.157, 0.50, 1.458) are for RC = 1 (acetonitrile). Atom type 100 is a dummy atom with $q = \sigma = \varepsilon = 0$.

This example does not have internal variations; however, for consistency the extra information can be left on the formally blank lines. In creating a new Z-matrix file, it is normal to edit an old one to get the formats right. More complex Z-matrix files have been provided in the test jobs. Since most user mistakes are made in the Z-matrix file, the program automatically checks for many potential problems. If any are found, they are noted in the ot file. When running long BOSS jobs, display the plt files regularly to check the solute structures. This can also reveal user errors; there are no known program errors.

A library of Z-matrices for the all-atom force field is provided in the subdirectory aazmat. These are ready for energy minimizations, as in the "optimize" test job.

10.13 Z-matrix Input for Custom Solvents

The Z-matrix input needed to designate a custom solvent molecule is nearly identical to that for the solutes. The only differences are: (1) the title is only five characters (A5) for proper printing purposes, (2) perturbations are not allowed so all final type entries should be zero or the same as initial type, and (3) the geometry variation and harmonic restraint sections should

have no entries, as shown below. An example Z-matrix for a fully flexible, all-atom methanol solvent is shown below. Note the way the methyl group is treated for concerted, yet flexible, rotation of the three hydrogens.

```
CH3OH
1 O      78     0    0    0.000000     0     0.000000    0     0.000000
2 DUM    -1     0    1    0.500000     0     0.000000    0     0.000000
3 DUM    -1     0    2    0.500000     1    90.000000    0     0.000000
4 H      79     0    1    0.945000     2    90.000000    3   180.000000
5 C     378     0    1    1.430000     4   108.500000    2   180.000000
6 HC    295     0    5    1.090000     1   109.500000    4   180.000000
7 HC    295     0    5    1.090000     1   109.500000    6   120.000000
8 HC    295     0    5    1.090000     1   109.500000    6   240.000000
                  Geometry Variations follow    (2I4, F12.6)
                  Variable Bonds follow         (I4 or I4-I4)
0004-0008
                  Additional Bonds follow       (2I4)
                  Harmonic Restraints follow    (2I4, 4F10.4)
                  Variable Bond Angles follow   (I4 or I4-I4)
0005-0008
                  Additional Bond Angles follow (3I4)
000100050007
000100050008
000700050008
                  Variable Dihedrals follow     (3I4,F12.6 or I4-I4)
000600580058  15.0
000700000000   5.0
000800000000   5.0
                  Additional Dihedrals follow   (6I4)
000400010005000700580058
000400010005000800580058
                  Domain Definitions follow     (4I4)
                  Final blank line
```

For a custom solvent, the programla 297.9001 Tm(generatles them)Tj 0.0551 Tw 12 0 0 12324.78a 297.900

)placings a custom

used. *E.g.*, for liquid benzene, a good choice would be to make atom NCENTS a dummy atom (type -1) at the ring center. ICUTAS should be used for large solvent molecules with the selected atoms wel distributed about the solvent molecule.

The AUTO feature for additional bond and dihedral angles has not been implemented for flexible solvent molecules. Any Z-matrix can be automatically expanded to include these entries by doing a single-point calculation with the xSPM script. Please see the MCliquid test job for examples and a tutorial.

11 PDB Input

Though rare, the BOSS program can read a PDB (Protein Data Bank) file to input the coordinates for the solute(s). Solvent can then be added, as usual, from the boxes or in file. The solute is treated as a rigid entity in this case, *e.g.*, no intramolecular degrees of freedom are

```
HETATM    6  C6  DMP     1      -2.514  -0.620  -0.067    0.200  3.800  0.170
HETATM    7  C7  DMP     1       2.514  -0.620   0.067    0.200  3.800  0.170
END
```

This allows facile input for a rigid solute, e.g., for linear response calculations. The above file is dmp.pdbp in the aazmat directory.

12 Mind Format and Reaction Path Following

The program is able to read a file containing a sequence of structures that are perturbed between in free energy calculations. This was used in the Jorgensen group to study reactions in solution for which a "movie" has been made using the reaction path following procedure in GAUSSIAN 92. See the following papers for applications to Diels-Alder and Claisen reactions: *J. Am. Chem. Soc.*, **113**, 7430 (1991); **114,** 10966 (1992). The structures that are generated along the minimum energy reaction path are placed sequentially in one file, which replaces the usual BOSS Z-matrix file. BOSS will then perturb between the structure entries that are requested using RC0, RC1, and RC2 to identify the reference and two perturbed structures. This then allows calculation of the change in free energy of solvation along the reaction path. The structures in the file are treated as one solute and are held rigid during the simulations. There is automatic processing in BOSS that (1) does a least-squares fit to maximally overlay the reference and two perturbed solutes, and (2) adds a dummy solute atom as the last atom in each structure. It is placed near the structures' centers and has the same coordinates in each structure. This dummy atom is set to be NROTA1, so the rigid-body rotations for the structures are made about this point.

For example, to begin a job to perturb the 4th structure in the file to the 3rd and 5th, one specifies in the command file:

```
time $boss 111 $configurations 4.0 3.0 5.0
```

ICALC is handled as usual, 1 for set-up, 0 to continue. MIND needs to be specified in the parameter file as the solute format, and it remains as MIND for all subsequent runs. The solvent origin can either be BOXES or IN. The latter uses the solvent coordinates in the in file from a previous run; it is often advantageous in this case to run the calculation for *ca.* 5000 configurations with only solute moves (NSCHG = 1) to help position the solute well in the solvent hole.

The format for the Mind coordinate file is given below. Free format is used and a pound sign (#) in column 1 for any line causes the line to be ignored, *e.g.*, a comment line. Note that the partial charge and Lennard-Jones parameters have to be specified for each atom at the end of its line.
Title line with a 30 character title beginning in column 21 (1-20 blank).
Then, coordinate lines for structure 1:

```
atomic # (integer), x, y, z, q, sigma, epsilon (real), atom symbol
(optional)
%     to end structure 1
Coordinate lines for structure 2
%     to end structure 2
Coordinate lines for structure 3
%     to end structure 3
etc.
```

The test job rxnmind provides an example of this procedure.

13 Pure Liquid Simulations

Monte Carlo simulations in the NVT and NPT ensembles can be performed for the eleven pure solvents that have been provided in the stored boxes and for custom solvents. A Z-matrix file for a dummy solute is needed with a single atom of type -1, *e.g.*,

```
Liquid Water Simulation
0001 DUM   -1
(then 10 blank lines)
```

The solvent name (SVMOD1) needs to be specified at the top of the parameter file and NSCHG can be set very large to not bother moving the dummy particle. Use ICUT = 0 and make WKC large, *e.g.*, 10,000. The atom type of -1 causes the dummy atom to be ignored in all energy calculations. The test job MCwater illustrates a MC simulation for pure TIP4P water. For custom solvents, the solvent Z-matrix file is needed (Section 10.13). The test job MCliquid illustrates MC simulation for a pure custom liquid, OPLS-AA ethanol, and other cases.

14 Cluster Simulations

Monte Carlo simulations can be performed for a solvent cluster with or without solutes. Periodic boundary conditions are <u>not</u> used. Since the volume is not well-defined, only NVT calculations are performed (no volume moves). The system is initially set up by designating in the parameter file the central solute atom for the cluster (ICAPAT; it may be a dummy atom), the initial radius of the cluster (CAPRAD in Å), and the restoring force constant (CAPFK in kcal/mol-Å^2). With the solvent origin as BOXES, the program automatically generates the solvent coordinates from the stored boxes, and discards any beyond CAPRAD or within 2.5 Å of a solute atom. NMOL, IBOX, and BCUT in the parameter file are ignored; the program will determine NMOL based on the choices of ICAPAT and CAPRAD. With the solvent origin as IN, the program will read the initial solvent coordinates from the in-file and center the cap on ICAPAT. If NMOL is less than the number of solvent molecules in the in-file, the remaining solvent molecules with the energetically worst interactions with the solute are discarded.

A half-harmonic potential is applied to keep solvent molecules from drifting away, if desired. The solute-solvent energies are incremented by CAPPOT, where R is the distance from atom 1 of the solvent molecule to ICAPAT:

$$\text{CAPPOT} = 0;\ R \leq \text{CAPRAD}$$

$$\text{CAPPOT} = \text{CAPFK}(R - \text{CAPRAD})^2;\ R > \text{CAPRAD}$$

For simulation of the true cluster, select CAPFK to be zero. In other applications, a weak restoring potential, *e.g.*, CAPFK = 0.5, may be desirable. If ICAPAT is a dummy atom alone in a separate solute (*e.g.*, solute 3), the program will not move it, allowing the cap center to be stationary.

In-files for several equilibrated water caps are provided in the solbox subdirectory, *e.g.*, wcapin.22 is a water sphere with 22 Å radius. These files can be used to obtain initial solvent coordinates for simulations of solutes in a water cap; for ICALC = 1, the solute and solvent origins would be ZMAT and IN. The 22 Å cap has 1503 water molecules; the 1503-NMOL waters with the worst interactions with the solute are automatically discarded. The test job MCdroplet illustrates the input for the simulation of a pure water cluster.

15 Solventless and Molecular Mechanics Calculations

15.1 Continuum Simulations

The program can also be used to perform Monte Carlo simulations for solute(s) in the absence of solvent. This is useful for obtaining gas-phase or continuum model reference data. In the parameter file, the solvent designation needs to be blank, *e.g.*, SVMOD1 = SVMOD2 = NONE, and the dielectric constant (ε) for the medium should be specified where indicated – the table below may be useful. The dielectric constant is used to scale all Coulombic interactions. The only other change is in the parameter file where the solvent box size IBOX, NMOL and NMOL2 need to be specified as zero.

The program will then perform NVT calculations as instructed in which every attempted move will be for the solute(s). The output only contains the thermodynamic results and dihedral angle distributions. Note that vacuum calculations could be used to search for conformational minima of a single solute or minima for solute pairs. High T runs could generate many starting structures which could then be quenched in low T simulations. As usual, use ICALC = 1 to begin the simulation and ICALC = 0 to continue it (Section 8). See test job MCgas for an example of a standard MC simulation of a single molecule in the gas phase.

Solvent	T (°C)	ε	bp (°C)
vacuum		1.0	
argon	–191	1.5	–185.9
water	25	78.3	100.0
methanol	25	32.66	64.5
ethanol	25	24.55	78.3
formamide	20	111.0	210.5
NMA	32	191.3	206.7
acetic acid	20	6.17	117.9
acetonitrile	25	35.94	81.6
DMSO	25	46.45	189.0
dichloromethane	25	8.93	39.6
chloroform	20	4.81	61.2
carbon tetrachloride	25	2.23	76.6
dimethyl ether	25	5.02	–24.8
diethyl ether	25	4.20	34.4
THF	25	7.58	66.0
benzene	25	2.27	80.1

Alternatively, **GB/SA** solvation can be invoked by declaring SVMOD1 = GBSA. The total energy for the sampling is then the gas-phase energy plus the free energy of hydration using the GB/SA model. The GB/SA treatment is based on the model described in Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C., *J. Phys. Chem. A*, **101**, 3005-3014 (1997). A principal difference is that OPLS-AA sigmas divided by two (σ/2) are used as the van der Waals radii, R_{vdW}, whereas Still et al. used OPLS-UA sigmas and a united-atom model for CH_n groups. Improved results are obtained with our OPLS-AA implementation; a paper summarizing results is in the bossman directory.

15.2 Energy Minimizations

More direct energy minimizations for the solute(s) are performed via ICALC = 2 and choosing an optimization method in the parameter file. As in continuum simulation, SVMOD1 and SVMOD2 must be set to NONE in the parameter file. The bond lengths, bond angles and dihedral angles that are declared as variable in the Z-matrix are optimized. The aazmat subdirectory contains all files needed to easily perform energy minimizations. The test job "optimize" also illustrates the standard setup for geometry optimizations with both the force field and QM methods. The provided optimizers do not use analytical first-derivatives of the energy. The following optimization methods are available in the program. **For single-point QM calculations, see QMNAME in Section 9.**

Non-Gradient Methods

- Downhill simplex method
- Powell's direction set method
- Hooke-Jeeves direct search method

Gradient Methods

- Davidon-Fletcher-Powell (Fletcher-Powell) method
- Broyden-Fletcher-Goldfarb-Shanno (BFGS) method
- Conjugate gradient method

Random Methods

- Simulated annealing

Both the non-gradient and gradient methods are designed to find a local minimum near the starting geometry. The gradient methods assume that the objective function is quadratic, and therefore they are not likely to be efficient and may result in a wrong solution if the starting geometry is far off the minimum. When an initial guess is poor, it is often advantageous to get a crude geometry using non-gradient methods and refine it through gradient methods. The simulated annealing method tries to find a global minimum rather than a local minimum. The basic principle behind this is essentially the same as the continuum simulation method, except that in the simulated annealing, the temperature decreases after a certain number of function evaluations and the system finally settles down at a minimum. Starting from an initial structure, the algorithm takes a step and the energy is evaluated. Any downhill moves are accepted. Some uphill moves are also accepted according to the Metropolis criteria, which allows the algorithm to jump over energy barriers. The process repeats from this new structure. As the optimization proceeds, the temperature decreases and the length of the steps declines as well. Thus the algorithm begins to concentrate more on a minimum and the optimization is terminated when the convergence criteria are met. Four parameters, NSAEVL, NSATR, SATEMP, and SART, can be set in the parameter file for the simulated annealing method. These are explained in the "Parameter File Input" section of this manual. This method may also be used to find a local minimum if the initial temperature is set very low, *e.g.*, SATEMP = 1.0E-10, and the temperature decreases fast, *e.g.*, NSATR = 0 and SART = 0.1. Since the simulated annealing method is usually slow, only a single pass is made for the geometry optimization. The user is expected to resubmit the resulting geometry as a new starting geometry for the next simulated annealing to check if this gives the same geometry or to begin a new search for the global minimum.

Since most methods listed above perform unconstrained optimizations, bond lengths and angles may become unreasonable if the initial guess is far off the minimum or the Z-matrix is poorly written, *e.g.*, negative bond lengths or bond angles greater than 180° may result. If this happens or bond angles approach 0° or 180°, error or warning messages are issued in the ot file and the optimization may be terminated. In this case, the Z-matrix must be rewritten in such a way that the linearity problem does not occur at any bond angles throughout the optimization (use dummy atoms, for example). For two solutes, the intersolute geometry normally requires 6 variables to be optimized. It is often convenient to use r, θ and ϕ for the first atom of the second solute, θ and ϕ for the second atom and ϕ for the third. The program knows to set the corresponding force constants to zero for any such variable bond lengths, bond angles and dihedral angles that involve atoms in two solutes; declare the type for any such dihedral angle as zero. These six variable declarations should be removed from the Z-matrix before Monte Carlo simulations are run. The program also knows to make force constants involving -1 dummy atoms zero, so dummies can be used in defining the optimized variables.

It may be worthwhile to try different optimizers to see if they converge to the same or different minima. If an optimization exceeds the maximum number of iterations, it can be restarted from the termination point. To do this, declare NEWZMAT in the parameter file, which causes the final Z-matrix to be written to the sum file. Delete the old Z-matrix file, copy the sum file to be the new Z-matrix file, and submit the job for another optimization cycle. It is recommended to make at least one such resubmission to verify convergence, particularly for relatively flat potential energy surfaces. Recall that FTOL (Section 9) controls the convergence criterion. More information on the optimization methods can be found in subroutines SIMPLEX, POWELL, FLEPOW, HOOKE, BFGS, CONGRD, and SIMANN.

In summary, for finding global minima, try simulated annealing or conformational searching (below), for poor initial geometries, try non-gradient methods, and for geometries already close to minima, try gradient methods. Currently, the maximum number of variables that can be optimized is 450 without redimensioning.

15.3 Dihedral Angle Driving

Repetitive energy minimizations can be performed as one dihedral angle is varied over 360°. Set ICALC = 3 and specify the atom number for the chosen dihedral angle as MXCON in the command file. The increment in degrees between minimizations needs to be specified as RC0 in the command file. The parameter file should be set up as for a single energy minimization discussed above. Energy minimizations will be performed every RC0 degrees from 0° through 360°. The first calculation is performed with the value of the dihedral angle set to zero; the value of the dihedral in the Z-matrix is ignored. *The chosen dihedral angle should be removed from the list of variable dihedral angles and added to the list of additional dihedral angles in the Z-matrix file to avoid optimizing it during each minimization*. The results for each minimization are written to the ot file; a summary is given at the end of the ot file and also in the sum file. This feature facilitates examining conformational energy profiles. The *dihxmgr* script that is provided in the dihdrive test job will create an xmgr file with the results for easy display.

15.4 Potential Surface Scanning

A potential surface can be scanned by repetitive energy minimizations as one variable bond length, bond angle or dihedral angle is varied. The chosen variable may be intra- or intermolecular. Set ICALC = 3 and specify the atom number for the chosen variable as

MXCON in the command file. The increment in Å or degrees between minimizations needs to be specified as RC0 in the command file. RC1 = 1.0, 2.0 or 3.0 to vary the bond length, bond angle or dihedral angle for atom MXCON. RC2 specifies the total number of energy minimizations to be performed. The first calculation uses the value of the variable in the Z-matrix. The value is incremented by RC0 for each subsequent calculation. The parameter file should be set up as for a single energy minimization discussed above. The results for each minimization are written to the ot file; a summary is given at the end of the ot file and also in the sum file. *Normally, the chosen variable should be removed from the list of variable bonds, bond angles or dihedral angles and added to the list of additional ones in the Z-matrix file to avoid optimizing it during each minimization*. See the scan test job.

15.5 Normal Coordinate Analysis

The aim of the normal coordinate computations in BOSS is to calculate frequencies (from eigenvalues) and nuclear displacements (from eigenvectors) for each normal mode of vibration from structural parameters, atomic masses and the OPLS force field. A normal coordinate is defined as a motion in which all atoms vibrate at the same frequency and phase, but indeterminate amplitudes. The normal coordinate computations can yield only a relative magnitude of the motion of each atom during a given normal mode. The two most common uses of the normal coordinate analysis are to characterize an optimized structure as a minimum or as a saddle point and to compare OPLS-derived frequencies with experimental values. A completely optimized ground-state structure with the OPLS force field should have no negative frequencies as long as the same force field is employed in the analysis. The theoretical background for normal coordinate analysis can be found in the following literature:

1. W. D. Gwinn, *J. Chem. Phys.* **55**, 477 (1971).
2. E. B. Wilson, Jr., J. C. Decius and P. C. Cross, Molecular Vibrations, McGraw-Hill, N.Y., 1975.
3. S. Califano, Vibrational States, John Wiley and Sons, London, 1976.
4. S. R. Niketic and K. Rasmussen, The Consistent Force Field: A Documentation, Springer-Verlag, 1977.
5. M. Diem, Introduction to Modern Molecular Vibrational Spectroscopy, John Wiley and Sons, N. Y., 1993.

The thermodynamic quantities, heat capacity, partition function and entropy are also calculated for translational, rotational and vibrational degrees of freedom for a given temperature after the normal coordinate computations.

Running Normal Coordinate Analysis. The NCA is carried out when NMODE is set to 1 in the parameter file and ICALC is 2, 3 or 4. If ICALC = 2 or 3, the geometry optimization is performed first, followed by the normal coordinate computation on the optimized geometry. If ICALC = 4, the NCA is performed without geometry optimization. The reference geometry (RC0) is always used for the analysis. Note that only the designated variable bonds, angles, and dihedral angles are optimized in BOSS geometry optimizations. Therefore, all other bonds, angles, and dihedral angles are treated as "rigid", *e.g.*, the first derivatives of the potential energy with respect to Cartesian coordinates associated with them are set to zero and the corresponding second derivatives are set to very large numbers. Then, normal modes corresponding to stretching, bending and torsion of these rigid parameters can be separated from the rest of the normal modes, *e.g.*, these vibrations will have very high frequencies.

Limits. The maximum number of atoms which can be handled in the normal coordinate analysis is currently 100 (MXFATM in the subroutine CALFREQ). This number has been set to a relatively small number since memory requirements for the second derivatives and normal coordinates are quite large. For molecules with more than 100 atoms, one will have to increase MXFATM and recompile.

Thermodynamic Calculations. At the end of normal coordinate computations, thermodynamic quantities using the resulting vibrational frequencies are obtained via statistical mechanics. Since the electronic energy at 0 K in the gas phase is not available, only zero-point vibrational energy, its thermal correction, and translational and rotational energies are reported. Vibrational, translational and rotational entropies and heat capacities are also computed. There are four configurable parameters unique to the thermochemical calculations: NSYM, NCONF, FSCALE and FRQCUT. Note that they specified are in the BOSS parameter file along with the temperature.

NSYM The symmetry number of the molecule.

Point Group	NSYM	Point Group	NSYM	Point Group	NSYM
C_1, C_i, C_s	1	D_2, D_{2d}, D_{2h}	4	$C_{\infty v}$	1
C_2, C_{2v}, C_{2h}	2	D_3, D_{3d}, D_{3h}	6	D_h	2
C_3, C_{3v}, C_{3h}	3	D_4, D_{4d}, D_{4h}	8	T, T_d	12
C_4, C_{4v}, C_{4h}	4	D_6, D_{6d}, D_{6h}	12	O_h	24
C_6, C_{6v}, C_{6h}	6	S_6	3		

NCONF The number of equivalent conformers. This is normally set to 1, the default value. However, for butane as an example, one could perform an NCA calculation with NCONF = 1 for the trans form and NCONF = 2 for the gauche form to obtain the correct entropy change (–R ln 2) reflecting the two equivalent gauche isomers.

FSCALE Frequency scale factor (default = 1.00). Vibrational frequencies are scaled with FSCALE for computing vibrational energies.

FRQCUT Cutoff frequency (default = 500 cm^{-1}). If scaled frequencies are less than or equal to FRQCUT, corresponding vibrations are treated as classical rigid rotations in computing the vibrational energy, *e.g.*, E = RT/2.

15.6 Conformational Searching

Conformers are geometries of a molecule which correspond to local minima on the potential energy surface. Conformational searching is a procedure which locates those minimum energy structures.[1] Usually, a search consists of three stages:

1. generation of starting geometries,

2. energy minimization to nearby minimum-energy conformers, and

3. identification of unique conformers.

(1) The first stage controls the overall effectiveness of the search in reaching convergence. A number of methods have been developed for the generation of starting geometries, which include systematic search, model building, random search, distance geometry, genetic algorithm, and molecular dynamics methods. BOSS implements an internal coordinate Monte Carlo method (ICALC = 5) and stochastic search (ICALC = 6).[2,3]

For the MC method, it is assumed that small structural changes in low-energy conformers favor the formation of new low-energy minima; thus, searching of the low-energy regions of conformational space is attempted by choosing starting geometries from among previously found low-energy conformers. This method's tendency to focus on certain regions of

When the two structures are superimposed, the RMS deviation for a least-squares overlay of all atoms except Hs on carbon must be smaller than CSRMS.

3. Internuclear distance criterion (CSRTOL).
   The sum of squared internuclear distances between all possible atom pairs excluding Hs on carbon must agree within CSRTOL.

Energy limits. For the storage of conformers, there are three user-configurable options:

- CSREUP (relative upper bound energy).
  At any point in a conformational search, only the conformers within CSREUP kcal/mol of the current lowest minimum (CLM) are stored for future use. The CLM is the lowest energy structure found during any incomplete conformational search. As a search proceeds, the CLM may become lower. In this case, some of previously stored conformers will be discarded, if they are beyond CSREUP of the new CLM.

- CSAELO (absolute lower bound energy).

- CSAEUP (absolute upper bound energy).
  A newly minimized conformer must have a total energy between CSAELO and CSAEUP to be stored.

Other Options. ICSOPT is an array of eight integers which are used to set some on/off options in the conformational searching. Since it has a format of `8I1` in the parameter file, in principle, eight on/off options can be set. So far, only the first digit is used, but the remaining seven digits may be used in the future. All digits are set to 0 by default.

- ICSOPT(1). Refine Geometry.
  If ICSOPT(1) = 1, the resulting conformers after a conformational search are reoptimized using the BFGS method with a higher precision (convergence criterion = 1.0E-6 kcal/mol). This may be useful when the optimization algorithm in the parameter file is a non-gradient method or FTOL is relatively high. The geometry refinement may be performed on the conformers stored in the up file. In this case, set both NEWRUN and MXCON to zero in the command file and set ICSOPT(1) to 1 in the parameter file, then run a conformational search.

Running a Conformational Search.

The xCS??? and xSS??? scripts are used for routine cases – see the boss/scripts directory.

(a) Monte Carlo conformational searching is carried out when ICALC = 5 and the internal coordinates to sample are listed at the end of the Z-matrix file or determined by the program (section 10.11). In the latter case, dihedral angles within rings are not sampled. The number of trial structures generated is designated by MXCON in the command file. A search may be restarted by setting NEWRUN to 0 in the command file. In this case, the up file must contain information on the conformers found in the previous search. Otherwise, a new search is initiated. Normal coordinate analysis is performed on each of the conformers at the end of the search, if NMODE is set to 1 in the parameter file. Conformers which have imaginary frequencies are not true minima. The options related to conformational search are CSETOL, CSRMS, CSRTOL, CSREUP, CSAELO, and CSAEUP, which have been described previously. The number of conformers produced depends mainly on the number of trial

structures and the above options. For a flat potential energy surface such as some intermolecular complexes, it may be a good idea to loosen the criteria for a unique conformer (*e.g.*, increase CSETOL, CSRMS, and/or CSRTOL, especially CSRTOL) in order to classify structurally very close conformers to the same conformer. The production of a new conformer becomes less frequent as the search proceeds. In order to make sure that convergence has been reached, set the number of trial structures large enough so that a new conformer is not found for a large number of steps at the end of the search. However, running the search for a longer period does not guarantee that either there will be no more new conformers or the global minimum has been found.

(b) Stochastic conformational searching is carried out when ICALC = 6. The input is the same as for the Monte Carlo search except the sampled internal coordinates do not have to be designated; all atoms are kicked on each trial. This is the preferred procedure when different ring conformers are desired. The line after the Conformational Search line in the Z-matrix can be blank or contain the following four variables in format 36X, I4, 3F10.4:

NFIX - The number of attempts made to fix quickly the bond lengths after kicking the structure. The default value is zero; values of ca. 200 can significantly speed up the optimization, though high-energy conformers are often lost.

RAD - This is the upper bound radius for the kick of any atom. Larger values lead to sampling more of the conformational space, but the optimization time may increase and sometimes optimizations may fail or lead to odd geometries. The default value is 2.7, though the user may wish to try values of 0.5 – 5.0 with the higher values for acyclic molecules.

PUSH and HWIDTH - These control the distribution of kick sizes that are attempted. The default values of 0.2 and 2.0 are adequate, but the user can explore these too.

The test job consearch illustrates the standard setups for a conformational search.

Printing of Results. The parameter IPRNT in the command file controls the verbosity of the output. It is normally set to 1, but if it is set to 3, the initial and final Z-matrices for each trial structure are printed. If the optimization algorithm is the simulated annealing method, intermediate step lengths are also printed. The ot file size may become very large. Therefore use IPRNT = 3 only for detailed clarification. The ot file contains any activities in each step of the search and the coordinates and Z-matrices of the final conformers. It may also contain normal coordinate analysis of conformers, depending on NMODE. The sum file contains a brief summary of the search. Cartesian coordinates of conformers are written to the plt file in MINDTOOL or PDB format for display. The save file contains full Z-matrices and total energies. The current status of the search is saved in the up file, which can be used for restarting a search later.

Limitations. The maximum number of internal coordinates (MAXCS) which can be sampled during a MC search is set to 450. The subroutine SEARCH uses the arrays NBOR, NBMOV and AC for neighbor list (NBLIST), pointer to neighbor list (NBPTR), and internal coordinates of conformers found (CSVAR), respectively, via FORTRAN's EQUIVALENCE statements. AC, ASOL, ASOL1, and ASOL2 are used for scratch space. If these arrays are used elsewhere during the execution of SEARCH, the real physical memory should be allocated in SEARCH. For this reason, the maximum number of conformers (MXCONF) which can be stored during the search is MXS*MXSAT*3/MXVAR, where;

MXS = maximum number of solvent molecules (2500).
MXSAT = maximum number of atoms in a solvent (50).

MXVAR = maximum number of variable bonds, angles, and dihedrals (450).

In the above case, MXCONF is 2500*50*3/450 = 833. The easiest way of increasing MXCONF is to reduce MXVAR in the subroutine SEARCH. MXVAR can not be larger than 450, unless NMAX is increased in all optimization subroutines.

References:

1. A. R. Leach "A Survey of Methods for Searching the Conformational Space of Small and Medium-Sized Molecules" in Reviews in Computational Chemistry , Ed. K. B. Lipkowitz and D. B. Boyd, Vol. 2, VCH, 1991.

2. G. Chang, W. C. Guida, and W. C. Still, *J. Am. Chem. Soc.* **111**, 4379 (1989).

3. M. Saunders, *J. Am. Chem. Soc.* **109**, 3150 (1987).

16 Available Solvent Boxes

The solvent boxes that have been provided with the program and references that describe the development of their potential functions are listed below. The boxes have been equilibrated at 25 °C except for dimethylether (-24.84 °C), propane (-42.07 °C), and argon (-185.85 °C) which were equilibrated at their normal (1 atm) boiling points.

Solvent	NMOL	Dimensions (Å)	References for Potential Functions
Water	216	18.6 X 18.6 X 18.6	*J. Chem. Phys.* **79**, 926 (1983) TIP3P, TIP4P
	250	17.1 X 17.1 X 25.6	*Mol. Phys.* **56**, 1381 (1985) TIP4P
	267	20.0 X 20.0 X 20.0	*J. Chem. Phys.* **112**, 8910 (2000) TIP5P
	324	18.6 X 18.6 X 27.9	
	400	20.0 X 20.0 X 30.0	
	512	25.0 X 25.0 X 25.0	
	750	25.0 X 25.0 X 37.5	
Methanol	128	20.9 X 20.9 X 20.9	*J. Phys. Chem.* **90**, 1276 (1986)
	192	20.9 X 20.9 X 31.3	
	267	26.7 X 26.7 X 26.7	
	400	26.7 X 26.7 X 40.0	
Acetonitrile	128	22.6 X 22.6 X 22.6	*Mol. Phys.* **63**, 547 (1988)
	192	22.6 X 22.6 X 33.9	
	267	28.9 X 28.9 X 28.9	
	400	28.9 X 28.9 X 43.3	
Dimethyl ether	125	23.8 X 23.8 X 23.8	*J. Comput. Chem.* **11**, 958 (1990)
	267	30.3 X 30.3 X 30.3	
	400	30.3 X 30.3 X 45.4	
Propane	128	25.7 X 25.7 X 25.7	*J. Am. Chem. Soc.* **106**, 6638 (1984)
	192	25.7 X 25.7 X 38.6	
	267	33.0 X 33.0 X 33.0	
	400	33.0 X 33.0 X 49.5	
Chloroform	128	25.7 X 25.7 X 25.7	*J. Phys. Chem.* **94**, 1683 (1990)

	192	25.7 X 25.7 X 38.6	
	267	33.0 X 33.0 X 33.0	
	400	33.0 X 33.0 X 49.5	
Methylene Chloride	128	24.3 X 24.3 X 24.3	*J. Am. Chem. Soc.* **116**, 3494 (1994)
	192	24.3 X 24.3 X 38.4	
	267	30.4 X 30.4 X 30.4	
	400	30.4 X 30.4 X 45.6	
Tetrahydrofuran	128	25.8 X 25.8 X 25.8	*J. Comput. Chem.* **11**, 958 (1990)
	192	25.8 X 25.8 X 38.7	
	267	32.5 X 32.5 X 32.5	
	400	32.5 X 32.5 X 48.8	
Argon	128	18.5 X 18.5 X 18.5	*Mol. Phys.* **24**, 1013 (1972)
	192	18.5 X 18.5 X 27.8	
	267	23.0 X 23.0 X 23.0	
	400	23.0 X 23.0 X 34.5	
Carbon Tetrachloride	128	27.6 X 27.6 X 27.6	*J. Am. Chem. Soc.* **114**, 7535 (1992)
	192	27.6 X 27.6 X 41.4	
	267	34.6 X 34.6 X 34.6	
	400	34.6 X 34.6 X 51.9	
Dimethyl Sulfoxide	128	24.5 X 24.5 X 24.5	Unpublished
	192	24.5 X 24.5 X 38.8	
	267	30.7 X 30.7 X 30.7	
	400	30.7 X 30.7 X 46.0	

Note: The length of the MM calculations is sensitive to the number of sites in the solvent molecules. For N sites, N^2 intersolvent distances must be calculated to evaluate a solvent-solvent interaction energy. To a first approximation, the length of the calculations is, consequently, proportional to N^2, *i.e.*, the computer time required for a simulation in THF is 25/9 = 2.8 times that for a simulation in methanol. United atom CH_n groups are used for all stored solvents.

17 Ewald Summation for Long-Range Electrostatics

When periodic boundary conditions are used, Ewald summation can be employed to calculate the exact electrostatic energy of an infinite lattice of identical copies of the simulation cell. This supresses artifacts resulting from the simple cutoff of the long-range electrostatic interactions. The flag EWALD has to be set to 1 in the par file (see *ionpar* in the ionwater test job).

In the Ewald method, the electrostatic energy of the infinite lattice is given by

$$E_{Coulomb} = \underbrace{\sum_{i<j}^{(>1\text{-}4)} \frac{q_i q_j}{\boldsymbol{r}_{ij}} erfc\left(\sqrt{\alpha}\,\boldsymbol{r}_{ij}\right)}_{\text{R-space sum}} + \underbrace{\frac{1}{2V}\sum_{\boldsymbol{k}\neq\boldsymbol{0}} \frac{4\pi}{k^2}\left|\rho(\boldsymbol{k})\right|^2 e^{-k^2/4\alpha}}_{\text{K-space sum}} - \underbrace{\sqrt{\frac{\alpha}{\pi}}\sum_{i=1}^{N} q_i^2}_{\text{Self-Interaction}}$$

$$\underbrace{- \sum_{i<j}^{(1-2,3,4)} \frac{q_i q_j}{\boldsymbol{r}_{ij}} erf\left(\sqrt{\alpha}\,\boldsymbol{r}_{ij}\right) + 0.5\sum_{i<j}^{(1-4)} \frac{q_i q_j}{\boldsymbol{r}_{ij}}}_{\text{Intrasolute nonbonded corrections}} \qquad \rho(\boldsymbol{k}) = \sum_j q_j\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}_j}$$

The infinite lattice is surrounded by a conducting medium with infinite dielectric constant ('tinfoil boundary conditions'). Otherwise an additional surface dipole term has to be added. Sufficiently large cutoffs for the Ewald r-sum and k-sum have to be used in order for the Ewald energy to converge. The r-space cutoff can be adjusted by RCUT. The r-sum is always converged by the program automatically setting the Ewald parameter α such that the value of the r-sum at RCUT is

$$\frac{erfc\left(\sqrt{\alpha}\,r_{CUT}\right)}{r_{CUT}} = 1.0\times10^{-6}$$

The k-space cutoff is set to $k_c = 7$. Therefore, allowed values of the r-sum cutoff are RCUT ≥ 9 Å. A cutoff of RCUT = 10 Å is recommended. For consistency, all other cutoffs, SCUT, CUTNB and ICUT should be set to the same value. Note that the Lennard-Jones interactions use the same cutoffs and are not affected by the Ewald summation. At the end of a run, the reported nonbonded component energy averages (EXX, ESX etc.) are averages of the r-sums only. Only the total nonbonded energy of the system can be averaged during a simulation with the Ewald summation.

18 Test Jobs

Prototypical test jobs have been provided in the *bosstst* directory. These can be mimicked for many new projects. In addition, the directory *aazmat* contains many all-atom Z-matrices that are ready for gas-phase optimizations, as described in the file INDEX. You may wish to perform some of these optimizations for your first use of BOSS. The file INDEX also gives instructions for performing a gas-phase MC simulation for any of these molecules. The *UNIX* command or *Windows* bat files for the test jobs are ready to be executed. If one is not using an x script from the *scripts* directory, all that is ever needed to begin work on a new problem is the linked program, and the command or bat, parameter, and Z-matrix files. For the typical initial set-up from the Z-matrix and stored solvent boxes, the in file is initially empty; otherwise it contains coordinates from a prior job. The files with the solvent boxes, watbox, org1box, org2box, and the files with the bond stretching and angle bending parameters, oplsua.sb and oplsaa.sb, must also be in the disk area or their location otherwise declared in the command or bat file. All other files are created by the program.

Each test job is ready to be executed. There is also a README file in each directory that should be read. Some of the test Monte Carlo runs are short, *i.e.*, just for instruction. In some cases previously generated ot files are included for comparisons. Also, view the generated plt files to see the systems in detail. Note: Due to arithmetic differences, results may vary slightly from computer to computer. Technically, for Monte Carlo simulations, the Markov chain diverges as soon as one configuration is accepted/rejected on one computer and not on the other.

Test Job	Type of Calculation
1 optimize	OPLS and QM energy minimizations for a molecule.
2 dihdrive	Compute energy as a function of a dihedral angle.
3 consearch	Conformational search for a molecule.
4 MCgas	MC simulation for an isolated molecule in the gas phase.
5 MCwater	MC simulation of liquid TIP4P water.
6 MCdroplet	MC simulation of a pure TIP4P water droplet.
7 MCliquid	MC simulation of a custom solvent: OPLS-AA ethanol.
8 ionwater	MC simulation of an ion in TIP4P water
9 linres	MC for a solute in water; obtain linear response energy components.
10 dihpmf	Compute pmf for rotation of a dihedral angle: dichloroethane in water.
11 rxnpmf	Compute pmf for a Michael addition in solution with AM1 energetics.
12 rxnmind	Compute a pmf for a stored reaction movie.
13 pairpmf	Compute pmfs for benzene dimer, methane dimer, and NaCl in water.
14 fepme	FEP calculations for mutating one molecule to another.
15 fepflex	FEP calculation for mutating flexible UA cyclohexane to THP in ethanol.
16 fepcrown	FEP calculation for mutating 18-crown-6/K^+ to 18-crown-6/Na^+.
17 fepktona	FEP calculation for mutating K^+ to Na^+ in methanol.
18 scan	Perform potential surface scans usually varying interatom distances.
19 SN2rxn	Compute pmf for an S_N2 reaction in solution.
20 MCGBSA	MC simulation for a solute with GB/SA solvation.
21 fepcharge	FEP calculation for perturbing initial charges for a molecule to new values.
22 ionnonaq	MC simulation of an ion in a stored non-aqueous solvent.

18.1 Test Job 1 — optimize

Scripts are used to energy-minimize a molecule. The example is for acetaminophen. It can be optimized with the OPLS-AA force field using xOPT or with PDDG/PM3 using xPDGOPT.

There are hundreds of Z-matrices for molecules and complexes in the aazmat directory that are ready for optimization. Hundreds of Z-matrices for drugs are also provided optionally in the drugs directory. See the file INDEX in both directories for a listing of the molecules.

18.2 Test Job 2 — dihdrive

This illustrates the performance of repetitive energy minimizations as a dihedral angle, the reaction coordinate, is varied. The example is for the C=C-N-H angle in acetaminophen, which is the dihedral angle for atom 15, the H on N. Note that this angle in the Z-matrix from test job 1 has been moved from the list of variable dihedral angles to the list of additional ones. Otherwise, for each increment of the reaction coordinate, the angle would be reoptimized.

Repetitive energy minimizations can also be performed for variation of a chosen bond length or bond angle; see Section 15.4.

18.3 Test Job 3 — consearch

Conformational searches are performed for the all-atom model of 1,2-ethanediol using the xCS and xSS scripts. The program finds 7 unique conformers, as summarized in the sum and out files. The lowest-energy forms are tgg^-, ggg^- , and gg^-g, as expected (see *J. Am. Chem. Soc.*, **116**, 3892 (1994)).

18.4 Test Job 4 — MCgas

This illustrates a Monte Carlo simulation for an isolated molecule in the gas phase. The example is for OPLS-AA ethanol at 25 °C. Such calculations can provide conformer populations in the gas phase. They also yield the average gas-phase energy, which is needed for computing the heat of vaporization in conjunction with a simulation of the corresponding pure liquid, as in test job 7.

18.5 Test Job 5 — MCwater

A full Monte Carlo simulation is performed for liquid water with the TIP4P model at 25 °C and 1 atm. The system is set up from the stored box of 267 water molecules. Although this box is equilibrated, 1M configurations of equilibration followed by 2M configurations of averaging are requested in the command or bat file. The individual runs are each 250K, so there are 12 runs in all. The parameter file requests output of the three radial distribution functions (OO, OH, and HH) and the energy and energy-pair distributions. From the final output in the t4pote2 or t4psum file, one can readily compute the density and heat of vaporization for comparison with the values reported in *Mol. Phys.* **56**, 1381 (1985): d = 0.997 g cm^{-3} and $\Delta H_{vap} = -E_{total}/267 + RT = 10.69$ kcal mol^{-1}. The complete job takes about 15 minutes on a 800 MHz Pentium III.

18.6 Test Job 6 — MCdroplet

This example illustrates a simulation of a pure water cluster. An initially spherical cluster of water molecules is requested with the designation of ICAPAT = 1 as the cap center in the parameter file. The requested radius is 12 Å (CAPRAD) and no restoring force is to be applied

(CAPFK = 0). The Z-matrix contains only a single dummy atom as for the simulation of any pure liquid. The program automatically generates a water cluster about the dummy atom that happens to contain 238 molecules within 12 Å. Five runs of 10,000 configurations each are requested in the command file and require 55 seconds to execute on an SGI R/5000. This system has been run for 5 x 10^6 configurations at 25 °C without any water molecules detaching themselves from the cluster.

18.7 Test Job 7 — MCliquid

A MC simulation for a custom, meaning non-stored, solvent is performed. The example is fully flexibile OPLS-AA ethanol at 25 °C and 1 atm. The Z-matrix for the solvent molecules is provided in the liqzmat file, as described in Section 10.13. The system with 267 monomers is built from scratch, then equilibrated for 2M configurations and averaged for 2M configurations. This takes ca 1 hour on a 2.4GHz Pentium or 7.4 hours on an SGI R5000. The results can be compared with those given in *J. Am. Chem. Soc.* **118**, 11225 (1996). The computed energies and volumes are a little high; the output for the incremental energies and volumes also indicate that the equilibration period should be extended. The calculation could be restarted (011) and averaged for a new 2M configurations. In order to compute the heat of vaporization, the gas-phase energy from test job 4 is needed:

$$\Delta H_{vap} = E_{gas} - E_{liquid} / 267 + RT$$

The subdirectory *methanol* contains extensive instructions for running new systems in the README file.

18.8 Test Job 8 — ionwater

This is a simple MC simulation for an ion in water. One can learn about the solvent structure/coordination about the ion in this manner. The system contains the ion and 500 TIP4P water molecules in a periodic cube. The setup is for methylammonium ion: however, one could replace the ionzmat file with the Z-matrix for another ion, e.g., one in the aazmat directory. 1M configurations of equilibration are followed by 2M configurations of averaging.

18.9 Test Job 9 — linres

This illustrates a full MC simulation for a solute in TIP4P water. The usual protocol for computing the energy components for a linear response calculation, as specified in the .bat and par files, is followed including use of 500 water molecules, ICUT = 5, 9 Å cutoffs, and 3M configurations or equilibration followed by 10M configurations of averaging. The final energy components, hydrogen-bond numbers, and solvent-accessible-surface-areas are summarized at the end of the sum file (lrsum). The full calculation takes ca. 2 hours on a 500 MHz PentiumIII or 4 hours on an SGI R5000.

The scripts xLRCMP and xLRAA are provided for easy execution of linear response calculations. They include the automatic prediction of molecular properties with the propCMP and propAA programs. See the README file.

References:

"Prediction of Properties from Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water", E. M. Duffy & W. L. Jorgensen, *J. Am. Chem. Soc.*, **122**, 2878-2888 (2000).

"Prediction of Drug Solubility from Monte Carlo Simulations", W. L. Jorgensen and E. M. Duffy, *Bioorg. Med. Chem. Lett*., **10**, 1155-1158 (2000).

18.10 Test Job 10 — dihpmf

The d010cmd or d010.bat file runs one window of a potential-of-mean force for rotation of the Cl-C-C-Cl dihedral angle in 1,2-dichloroethane. The angle is perturbed from 10° to 0° and 20°. The 1M configurations of equilibration and 2M of averaging require 1.3 hours on an R5000 SGI. Note that the initial and final values of the dihedral angle are given as 0.0 and 1.0 in the Z-matrix file, pmfzmat. The cmd files are then set-up so that all that needs to be specified is the value of the Cl-C-C-Cl dihedral angle in degrees for the window. The ultimate purpose is to obtain the *gauche* and *trans* conformer populations, which are significantly affected by solvation (see *JACS* **117**, 11809 (1995)). The complete pmf (0° to 180°) can be obtained by submitting a single job as in the fullpmf subdirectory.

18.11 Test Job 11 - rxnpmf

This illustrates a pmf calculation for a reaction. A QM/MM calculation is performed with complete solute flexibility. The reaction is the Michael addition of methane thiolate to methyl crotonate. The reaction coordinate is taken as the C_β - S distance. The solute energetics are evaluated with AM1, and the solute-solvent interactions use the unscaled CM1A charges and generic OPLS-AA Lennard-Jones parameters for the solute atoms. The 390 solvent molecules come from the stored box of 400 OPLS-UA methanol molecules.

Only one window is executed with the r1850cmd or r1850.bat file corresponding to perturbations of the reaction coordinate from 1.85 Å to 1.775 and 1.925 Å. Only 400K configurations of equilibration and 400K of averaging are executed for illustration. This takes 53 min on an R5000 SGI. Other windows can be run in parallel or chained by straightforward creation of corresponding cmd/bat files.

18.12 Test Job 12 - rxnmind

This illustrates a pmf calculation for a reaction using Mind input. The reaction is the addition of *t*-butyl cation to isobutylene in THF. The reaction path came from RHF/6-31G* calculations and is stored as a 53 frame movie in the Z-matrix file. In this format, internal degrees of freedom for the solutes are not sampled and the non-bonded parameters for the solute atoms are included with each atom entry in the Z-matrix file. The choice of frames for the perturbations is given in the cmd/bat file. See: *J. Am. Chem. Soc.* **119**, 10846 (1997).

18.13 Test Job 13 - pairpmf

This illustrates computation of a pmf for separation of benzene dimer in water. One window is run for the ring center - ring center distance changing form 5.0 Å to 4.9 and 5.1 Å. The calculation was first performed in *J. Am. Chem. Soc.* **112**, 4768 (1990).

A subdirectory contains the files needed to determine the entire pmf for separation of two methanes in water via one job submission.

Subdirectory nacl is set up for running a pmf for separating Na^+ and Cl^- in water.

18.14 Test Job 14 - fepme

This illustrates the classic free energy perturbation of OPLS-UA methanol to ethane in TIP4P water, as in *J. Chem. Phys.* **83**, 3050 (1985). The entire FEP is run by submission of just the fepcmd file. Five windows are executed with $\Delta\lambda = \pm 0.10$. Each window covers 1M configurations of equilibration and 2M configurations of averaging. There are 200 water molecules and the solute in the periodic cube. The calculation requires 4 hours on an R5000 SGI or less than 1 hour on a 1 GHz Pentium. The original calculations took about a month on a Harris 80.

Subdirectories contain standard current setups for running FEP calculations in which one molecule is mutated into another in many solvents and in the gas phase. Subdirectory *ch3cf3* illustrates a complete FEP calculation for computing the difference in free energies of hydration for toluene and trifluoromethylbenzene.

18.15 Test Job 15 — fepflex

This job illustrates (a) a slightly more complicated mutation with changes in internal variables and (b) the use of a custom ("OTHER") solvent. The setup is for mutating united-atom cyclohexane (RC0 = 0) to tetrahydropyran (RC0 = 1) in UA ethanol where the first perturbation (window) is from RC0 = 0.125 to RC1 = 0.00 and RC2 = 0.25. Six runs of 100K configurations each are executed, though for the real calculation one would normally equilibrate for *ca.* 1M configurations and average for *ca.* 3M in each window using runs of *ca.* 250000. Three more windows are executed by chaining at RC0 = 0.375, 0.625, and 0.875. The total time needed on an R5000 SGI is 1.1 hours.

In the Z-matrix file, all bond lengths, angles and dihedral angles are designated as variable. The ring closure bond is between atoms 3 and 8; its stretching energy is included. Thus, the ring is fully flexible. The types for four dihedral angles change from united-atom type 01 to 41 or 44 as the alkane becomes the ether. The changes in types for the bond stretching and angle bending are handled automatically by the program. The computed change in free energy includes the changes in internal energy. Typically another mutation of cyclohexane to THP would be performed, *e.g.*, in the gas phase or bound to a receptor, to complete a thermodynamic cycle. For the custom solvent, the Z-matrix is given in the file, etoh.zmat. 260 molecules from the 267 box of equilibrated ethanol (solvent origin = IN) are requested in the par file. For more information on fep calculations with flexible solutes, please read flexpert.note in the note directory. Also, chexol2.z in aazmat provides an example of fully flexible OPLS-AA 1,2-cyclohexanediol going to cyclohexanol.

18.16 Test Job 16 — fepcrown

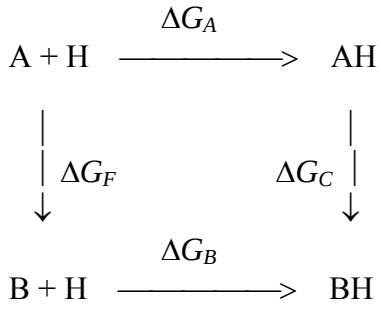
This is a perturbation of potassium ion to sodium ion bound to 18-crown-6. The solute is represented with the OPLS-AA force field. The solvent is UA methanol; 255 molecules are used from the stored box of 267. Only two windows are run with $\Delta\lambda = \pm 0.25$ for this small change. For illustration, the windows are equilibrated for 1M configurations and then averaged for 2 M. This takes 3.4 hours on an R5000 SGI. For proper sampling, at least 2 M configurations of equilibration and 10 M of averaging are recommended.

18.17 Test Job 17 — fepktona

This is the companion calculation for the fepcrown test job. A single potassium ion is perturbed to sodium ion in methanol. For illustration, the windows are again equilibrated for 1M configurations and then averaged for 2 M. This takes 1.6 hours on an R5000 SGI The difference in the free energy changes for the bound and unbound processes gives the relative

free energy of binding of the two ions in methanol. For the scheme below, H is the crown ether, A is K^+ and B is Na^+.

The experimental difference, $\Delta\Delta G_b$ ($K^+ \rightarrow Na^+$), is 2.45 kcal/mol. The value computed here is 3.40 ± 0.34 kcal/mol. The difference in free energies of solvation, ΔG_F, for the two ions in methanol is also known experimentally to be 17.3 kcal/mol. The value obtained in this test job is 17.11 ± 0.19 kcal/mol. See *J. Am. Chem. Soc.* **112**, 4411 (1990) for a closely related MD study.



$$\Delta\Delta G_b = \Delta G_B - \Delta G_A = \Delta G_C - \Delta G_F$$

18.18 Test Job 18 — scan

This illustrates potential surface scanning including an example of finding a transition structure for an S_N2 reaction. Other examples include computing potentials of mean force with GB/SA solvation.

18.19 Test Job 19 — SN2rxn

QM/MM MC/FEP calculations are used to compute a free energy of activation for an S_N2 reaction in solution. Linux only.

18.20 Test Job 20 — MCGBSA

This is similar to the MCgas test job with the addition of GB/SA solvation.

18.21 Test Job 21 — fepcharge

FEP calculations for computing changes in atom types both in the gas phase and aqueous solution. This allows determination of potential function changes on the free energy of hydration.

18.22 Test Job 22 — ionnonaq

This is similar to the ionwater test job except a non-aqueous solvent is used. The ion is specified in *ionzmat* and the solvent in *ionpar*. Again, one can learn about the solvent structure/coordination about the ion in this manner.

19 Output

The output for Monte Carlo simulations in the "ot file" is in two main sections. Before the Monte Carlo calculation begins or resumes, the Z-matrix is listed along with the potential function and torsional parameters. The total energy and its components are recomputed from the coordinates in the in file or computed for the first time if ICALC = 1 or 2.

If it is a continuation run (ICALC = 0), then it is required that OLD E = NEW E at the *beginning* of the new run within the limits of precision of the computer (at least 7 figures). If this does not happen, the user needs to understand why - investigate!

Many other parameters for the simulation are listed along with the energies. The values of RC0, RC1 and RC2 are then noted, followed by the coordinates for the reference solute and, if a perturbation calculation is being performed, for the two perturbed solutes - note that each of these two solutes may have multiple molecular fragments (sub-solutes). The Monte Carlo calculation then begins. When it finishes, the acceptance rate information for volume moves is printed. If a *ca.* 40% acceptance rate is not obtained, VDEL should be adjusted in the parameter file. Then, the following output is provided for Monte Carlo runs:

1. plots showing the history for several variable dihedral angles in the solutes,

2. the final total energy and its components along with the parameters for the simulation as above,

3. the final coordinates, solvent accessible surface area and volume for the solutes,

4. the averages for the thermodynamic properties including the two free energies which are repeated in fuller form below, average numbers of solute-solvent hydrogen bonds,

5. the solute-solvent atom-atom radial distribution functions and their integrals (coordination numbers) that have been requested in the parameter file,

6. the solvent-solvent and solute-solvent total energy and energy pair distribution functions,

7. the distribution functions for the variable dihedral angles,

8. the record of attempted and accepted moves for each solvent and solute molecule, and

9. the full report on the computed thermodynamic results including the averages for each run, the total averages and the standard deviations (1 σ) calculated from the fluctuations in the averages for each run. This output is provided when more than one run has been completed.

Note: results are included on the ΔH and ΔS for the two ΔG's that are computed in perturbation calculations. These are computed by an umbrella sampling procedure. Unfortunately, the statistical noise is so large that the computed ΔH's and ΔS's are usually too imprecise to be useful.

It should be noted that the average of a property over the entire simulation may or may not be equal to the average of the averages for each run (block). This equality holds for properties given by linear functions such as the total energy, energy components, and volume. However, it

does not hold for properties given by non-linear expressions including the fluctuation properties (heat capacity, coefficient of thermal expansion, and compressibility) and the free energy changes from the Zwanzig equation. In the latter case,

$$\Delta G_{avg} = -kT \ ln \ \{1/NRUN * [\sum_{i} \exp(-\Delta G_i / kT)]\}$$

where the summation is over i = 1 to NRUN, the number of blocks.

When the command file requests multiple runs on one job submission, a common practice is to name the resulting multiple ot files xxxota, xxxotb, xxxotc, xxxotd, etc. where xxx is the 3 letter identifier for the project.

Abbreviated output is provided for solute optimizations (ICALC = 2). Definitions for some of the output quantities are given below. All energies are in kcal/mol, all distances in Å, and all angles in degrees.

19.1 Some Variable Definitions

NMOL	The number of solvent molecules.
EDGE	The dimensions of the simulation cell.
RCUT	The solvent-solvent cutoff distance.
SCUT	The solute-solvent cutoff distance.
ICUT	The cutoff procedure for solute-solvent interactions.
NACCPT	The total number of Monte Carlo steps (configurations).
NRJECT	The number of attempted steps that were rejected (this should be approximately 60 % NACCPT).
MXCON	The number of steps requested for the current run.
NCON	The current step number for this run.
T	The temperature in °C.
P	The external pressure in atm.
RDEL	The range for solvent translations.
RDEL1	The range for translations of solute 1.
RDEL2	The range for translations of solute 2.
ADEL	The range for solvent rotations.
ADELS1	The range for rotations of solute 1.
ADELS2	The range for rotations of solute 2.
OLD E	The total energy for the last accepted configuration.
NEW E	The total energy for the last attempted configuration.
VOLD	The total volume in $Å^3$ for the last accepted configuration.
VNEW	The total volume for the last attempted configuration.
NSCHG	Frequency of attempted solute moves in configurations.
NVCHG	Frequency of attempted volume changes.
DIELEC	The dielectric constant.
ESOLD	The old (last accepted) total solute-solvent energy.
ESONE	The new (last attempted) total solute-solvent energy.
ESOL1	The same for the two perturbed solutes corresponding to RC1 and RC2.
ESOL2	
ESON1	
ESON2	

EXXOLD	The old solute-solute energy.
EXXNEW	The new solute-solute energy.
EXXOL1	The same for the perturbed solutes.
EXXOL2	
EXXNE1	
EXXNE2	
EDHIOL	The old torsional energy for the solutes.
EDHINE	The new torsional energy for the solutes.
ENBOL	The old intrasolute non-bonded interaction energy.
ENBNE	The new intrasolute non-bonded interaction energy.
ENBOL1	The same for the perturbed solutes.
ENBOL2	
ENBNE1	
ENBNE2	
EBCOLD	The old energy for the distance constraints.
EBCNEW	The new energy for the distance constraints.
EBCOL1	The same for the perturbed solutes.
EBCOL2	
EBCNE1	
EBCNE2	
EQMetc.	The total QM energy for the various solutes.
CUTOFF E	Solvent-solvent (non-aqueous only) energy correction for Lennard-Jones interactions neglected beyond the cutoff RCUT.

20 Contents of the Distribution Files

File Name	Description
In the boss or miscexec directory:	
BOSS	BOSS executable code.
pargen	Executable code for the par file generator.
PROPCMP	Executable code for properties predictions using CM1P charges.
PROPAA	Executable code for properties predictions using OPLS-AA charges.
bossinfo	awk script to make plots from ot files for ChemEdit display.
bar	Executable to generate xmgr input files from an av file.
oplsua.par	The latest version of the OPLS united-atom parameters.
oplsua.sb	Bond stretching and angle bending parameters (mostly from AMBER).
oplsaa.par	The latest version of the OPLS all-atom parameters.
oplsaa.sb	All-atom stretching and bending parameters.
watbox	File of standard water boxes.
org1box	File 1 of standard non-aqueous solvent boxes.
org2box	File 2 of standard non-aqueous solvent boxes.
suboss	Alternative style of command file for UNIX.
In the optional boss/source directory:	
boss.f	FORTRAN source code for BOSS.
pargen.f	FORTRAN source code for par file generator.
qmcalc.f	FORTRAN source code for the QM calculations.
ewald.f	FORTRAN source code for the Ewald calculations.
bar.f	FORTRAN source code for bar.

propcmp.f	FORTRAN source code for PROPCMP.
propAA.f	FORTRAN source code for PROPAA.
PARAM.i	AM1,PM3, and PDDG parameters.
PARAMS.i	CM solvation model parameters.
Makefile	Contains compiling macros. Please read.

The solbox subdirectory contains the following files and others for custom solvents and water clusters:

etoh267.in	Equilibrated in-file for 267 ethanol molecules (united atom).
etoh.zmat	Corresponding solvent Z-matrix file.
nmp267.in	Equilibrated in-file for 267 1-methyl-2-pyrrolidinone molecules.
nmp.zmat	Corresponding solvent Z-matrix file (united atom).
hex267.in	Equilibrated in-file for 267 hexane molecules (united atom).
hex.zmat	Corresponding solvent Z-matrix file.
wcapin.18	Equilibrated in-file for a water cluster with 18 Å radius.
wcapin.20	Equilibrated in-file for a water cluster with 20 Å radius.
wcapin.22	Equilibrated in-file for a water cluster with 22 Å radius.

The *aazmat* directory contains many Z-matrices for the OPLS all-atom force field. The *peptide* directory contains optimized Z-matrices for a low-energy structure of each common amino acid as a dipeptide using the OPLS-AA force field. The optional *drugs* directory contains Z-matrices for hundreds of common drugs. The *notes* subdirectory contains miscellaneous information on BOSS and MC simulations in general. The *bosstst* directory contains a subdirectory for each test job.

21 Appendix – No. of Molecules for Binary Solvent Mixtures

For two solvents x and y in a volume x: volume y mixture (V_x/V_y), the mole ratio is:

$$n_x / n_y = V_x d_x MW_y / V_y d_y MW_x$$

Then, for N total molecules,

$$N_x = N(n_x / n_y)/(n_x / n_y + 1)$$
$$N_y = N - N_x$$

Solvent	density, d (25 C)	MW
Water	0.997	18.02
Methanol	0.786	32.04
Ethanol	0.785	46.07
t-butanol	0.781	74.12
Hexane	0.6548	86.18
Acetone	0.784	58.08
Acetic acid	1.044	60.05
Ethyl ether	0.708	74.12
THF	0.884	72.11

E.g., for 80% aqueous acetone (80:20 v:v acetone:water) with N = 500 molecules,

$$n_{acetone} / n_{water} = 80x0.784x18.02 / 20x0.997x58.08 = 0.9759$$
$$N_{acetone} = 500x0.9759 / 1.9759 = 247$$
$$N_{water} = 500 - 247 = 253$$

22 Appendix - OPLS All-Atom Parameters

Parameters for the OPLS All-Atom Force Field

The following tables contain **some** published non-bonded and torsional parameters for the OPLS-AA force field. These, parameters for many other systems including heterocycles, nucleic acids and carbohydrates, and the bond-stretching and angle-bending parameters are distributed with the BOSS program in the files *oplsaa.par* and *oplsaa.sb*.

Form of the Force Field

Bond stretching: $$E_{bond} = \sum_{bonds} K_r (r - r_{eq})^2$$

Angle bending: $$E_{angle} = \sum_{angles} K_\theta (\theta - \theta_{eq})^2$$

Torsion: $$E_{dih} = V_1(1+\cos\phi)/2 + V_2(1-\cos 2\phi)/2 + V_3(1+\cos 3\phi)/2 + V_4(1-\cos 4\phi)/2$$

Non-bonded: $$E_{ab} = \sum_{i}^{on\ a} \sum_{j}^{on\ b} [q_i q_j e^2 / r_{ij} + 4\varepsilon_{ij}(\sigma_{ij}^{12}/r_{ij}^{12} - \sigma_{ij}^{6}/r_{ij}^{6})] f_{ij}$$

f_{ij} = 0.5 if i, j are 1,4; otherwise, f_{ij} = 1.0

References

W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. 118, 11225-11236 (1996).
W. Damm, A. Frontera, J. Tirado-Rives, and W. L. Jorgensen, J. Comput. Chem. 18, 1955-1970 (1997).
W. L. Jorgensen and N. A. McDonald, Theochem 424, 145-155 (1998).
W. L. Jorgensen and N. A. McDonald, J. Phys. Chem. B 102, 8049-8059 (1998).
R. C. Rizzo and W. L. Jorgensen, J. Am. Chem. Soc. 121, 4827-4836 (1999).
E. K. Watkins and W. L. Jorgensen, J. Phys. Chem. A 105, 4118-4125 (2001).
M. L. P. Price, D. Ostrovsky, and W. L. Jorgensen, J. Comput. Chem., 22, 1340-1352 (2001).

Table 1. OPLS-AA Non-Bonded Parameters for Hydrocarbons and Alcohols

atom or group	q, e^-	σ, Å	ε, kcal mol^{-1}
C, CH_4	-0.240	3.500	0.066
C, RCH_3	-0.180	3.500	0.066
C, R_2CH_2	-0.120	3.500	0.066
C, R_3CH	-0.060	3.500	0.066
C, R_4C	0.000	3.500	0.066
H, RH, alkanes	0.060	2.500	0.030
C, Benzene	-0.115	3.550	0.070
H, Benzene	0.115	2.420	0.030
C, CH_3 of toluene	-0.065	3.500	0.066
C, CH_2 of ethyl benzene	-0.005	3.500	0.066
C, $R_2C=$	0.000	3.550	0.076
C, RHC=	-0.115	3.550	0.076
C, $H_2C=$	-0.230	3.550	0.076
H, HC=	0.115	2.420	0.030
O, ROH	-0.683	3.120	0.170
H(O), ROH	0.418	0.000	0.000
H(C), CH_3OH	0.040	2.500	0.030
C, CH_3OH and RCH_2OH	0.145	3.500	0.066
C, R_2CHOH	0.205	3.500	0.066
C, R_3COH	0.265	3.500	0.066
C, COH phenol	0.150	3.550	0.070
O, phenol	-0.585	3.070	0.170
H, phenol	0.435	0.000	0.000

Table 2. OPLS-AA Non-Bonded Parameters for Sulfur Compounds and Amines

atom or group	q, e^-	σ, Å	ε, kcal mol^{-1}
S, RSH	-0.435	3.550	0.250
H(S), RSH	0.255	0.000	0.000
C, CH_3SH	0.000	3.500	0.066
C, RCH_2SH	0.060	3.500	0.066
C, R_2CHSH	0.120	3.500	0.066
C, R_3CSH	0.180	3.500	0.066
S, RSR	-0.435	3.550	0.250
C, CH_3SR	0.0375	3.500	0.066
C, RCH_2SR	0.0975	3.500	0.066
C, R_2CHSR	0.1575	3.500	0.066
C, R_3CSR	0.2175	3.500	0.066
S, RSSR	-0.2175	3.550	0.250
C, CH_3SSR	0.0375	3.500	0.066
C, RCH_2SSR	0.0975	3.500	0.066
C, R_2CHSSR	0.1575	3.500	0.066
C, R_3CSSR	0.2175	3.500	0.066
N, RNH_2	-0.900	3.250	0.170
H, RNH_2	0.350	0.000	0.000
C, CH_3NH_2	0.020	3.500	0.066
C, RCH_2NH_2	0.080	3.500	0.066
C, R_2CHNH_2	0.140	3.500	0.066
C, R_3CNH2	0.200	3.500	0.066

Table 3. OPLS-AA Non-Bonded Parameters for Ammonium Ions, Imidazoles, and Carboxylate Ions[a]

atom or group	q, e^-	σ, Å	ε, kcal mol^{-1}
N, RNH_3^+	-0.300	3.250	0.170
H, RNH_3^+	0.330	0.000	0.000
C, $CH_3NH_3^+$	0.130	3.500	0.066
C, $RCH_2NH_3^+$	0.190	3.500	0.066
C, $R_2CHNH_3^+$	0.250	3.500	0.066
C, $R_3CNH_3^+$	0.310	3.500	0.066
C, $C_{\varepsilon 1}$ in HID, HIE	0.295	3.550	0.070
C, $C_{\delta 2}$ in HID, C_γ in HIE	-0.015	3.550	0.070
C, C_γ in HID, $C_{\delta 2}$ in HIE	0.015	3.550	0.070
C, $C_{\varepsilon 1}$ in HIP	0.385	3.550	0.070
C, C_γ, $C_{\delta 2}$ in HIP	0.215	3.550	0.070
H, H on C_δ or C_ε in HID, HIE, HIP	0.115	2.420	0.030
N, N_δ in HID or N_ε in HIE	-0.570	3.250	0.170
H, H(N_δ) in HID or H(N_ε) in HIE	0.420	0.000	0.000
N, N_ε in HID or N_δ in HIE	-0.490	3.250	0.170
N, in HIP	-0.540	3.250	0.170
H(N), in HIP	0.460	0.000	0.000
C, CH_3- in 5-methylimidazole	-0.065	3.500	0.066
C, RCH_2- in 5-ethylimidazole	-0.005	3.500	0.066
C, $RCOO^-$	0.700	3.750	0.105
O, $RCOO^-$	-0.800	2.960	0.210
C, CH_3COO^-	-0.280	3.500	0.066
C, RCH_2COO^-	-0.220	3.500	0.066
C, R_2CHCOO^-	-0.160	3.500	0.066
C, R_3CCOO^-	-0.100	3.500	0.066

[a] HID, HIE, and HIP refer to unprotonated histidine (imidazole) with hydrogens on N_δ or N_ε, and protonated histidine (imidazole), respectively.

Table 4. OPLS-AA Non-Bonded Parameters for Guanidinium Ions, Tryptophan, and Amides[a]

atom or group	q, e^-	σ, Å	ε, kcal mol^{-1}
N, NH_2 in guanidinium	-0.800	3.250	0.170
H, NH_2 in guanidinium	0.460	0.000	0.000
C, guanidinium	0.640	2.250	0.050
N, NHR in alkylguanidinium	-0.700	3.250	0.170
H, NHR in alkylguanidinium	0.440	0.000	0.000
C, CH_3 in methylguanidinium	0.200	3.500	0.066
C, CH_3 in ethylguanidinium (EG)	-0.110	3.500	0.066
C, CH_2(δ) in Arg, EG	0.190	3.500	0.066
C, CH_2(γ) in Arg	-0.050	3.500	0.066
N, N_ε in Trp	-0.570	3.250	0.170
H, H_ε(N) in Trp	0.420	0.000	0.000
C, C_γ in Trp	0.075	3.550	0.070
C, $CH_{\delta,\varepsilon,\zeta,\eta}$ in Trp	-0.115	3.550	0.070
H, $H_{\delta,\varepsilon,\zeta,\eta}$ (C) in Trp	0.115	2.420	0.030
C, C_δ in Trp	-0.055	3.750	0.145
C, C_ε in Trp	0.130	3.750	0.145
C, C=O in amide	0.500	3.750	0.105
O, C=O in amide	-0.500	2.960	0.210
H, HCONRR’	0.000	2.420	0.015
N, 1° amide	-0.760	3.250	0.170
N, 2° amide	-0.500	3.250	0.170
N, 3° amide	-0.140	3.250	0.170
H(N), 1° amide	0.380	0.000	0.000
H(N), 2° amide	0.300	0.000	0.000
C, CH_3N- 2° amide	0.020	3.500	0.066
C, CH_3N- 3° amide	-0.110	3.500	0.066
C, R CH_2N - 3° amide (Cδ in Pro)	-0.050	3.500	0.066
C, R $_2$CHN - 3° amide (Cα in Pro)	0.010	3.500	0.066
C, CH_2(α) in Gly	0.080	3.500	0.066
C, CHR(α) in Ala	0.140	3.500	0.066
C, CRR’(α) in Aib	0.200	3.500	0.066

[a] R in RCONR'R" is neutral and uses alkane parameters.

Table 5. OPLS-AA Non-Bonded Parameters for Ethers, Acetals, Aldehydes, Ketones, and Carboxylic Acids.

Atom or group	q, e^-	σ, Å	ε, kcal mol^{-1}
O, ROR	-0.400	2.900	0.140
C, CH_3OR	0.110	3.500	0.066
C, RCH_2OR	0.140	3.500	0.066
C, R_2CHOR	0.170	3.500	0.066
C, R_3COR	0.200	3.500	0.066
H, CH_nOR	0.030	2.500	0.030
O, acetal	-0.400	2.900	0.140
C, $ROCH_2OR$	0.200	3.500	0.066
H, $ROCH_2OR$	0.100	2.500	0.030
C, ROCHROR	0.300	3.500	0.066
H, ROCHROR	0.100	2.500	0.030
C, $ROCR_2OR$	0.400	3.500	0.066
C, RCHO	0.450	3.750	0.105
O, RCHO	-0.450	2.960	0.210
H, RCHO	0.000	2.420	0.015
C, R_2CO	0.470	3.750	0.105
O, R_2CO	-0.470	2.960	0.210
H, CH_nCOR[a]	0.060	2.420	0.015
C, RCOOH	0.520	3.750	0.105
O(C), RCOOH	-0.440	2.960	0.210
O(H), RCOOH	-0.530	3.000	0.170
H, RCOOH	0.450	0.000	0.000

[a] H on alpha C of aldehyde and ketone. Alpha C uses alkyl C parameters (Table 1).

Table 6. OPLS-AA Fourier Coefficients (kcal/mol) for Torsional Energy Functions

System	Dihedral	V_1	V_2	V_3
alkane	H-C-C-H*	0.000	0.000	0.300
*changed 11/99	H-C-C-C*	0.000	0.000	0.300
	C-C-C-C*	1.300	-0.050	0.200
alkene	H-C-C=C	0.000	0.000	-0.372
ethylbenzene	H-C-C_{ar}-C_{ar}	0.000	0.000	0.000
	C-C-C_{ar}-C_{ar}	0.000	0.000	0.000
	H-C-C-C_{ar}	0.000	0.000	0.462
alcohol	H-C-O-H	0.000	0.000	0.450
	C-C-O-H	-0.356	-0.174	0.492
	H-C-C-O	0.000	0.000	0.468
	C-C-C-O	1.711	-0.500	0.663
phenol	H-O-C_{ar}-C_{ar}	0.000	1.682	0.000
thiol	H-C-S-H	0.000	0.000	0.451
	C-C-S-H	-0.759	-0.282	0.603
	H-C-C-S	0.000	0.000	0.452
	C-C-C-S	1.876	0.000	0.000
sulfide	H-C-S-C	0.000	0.000	0.647
	C-C-C-S	2.619	-0.620	0.258
	C-C-S-C	0.925	-0.576	0.677
disulfide	C-S-S-C	0.000	-7.414	1.705
	H-C-S-S	0.000	0.000	0.558
	C-C-S-S	1.941	-0.836	0.935
1° amine	H-C-N-H	0.000	0.000	0.400
	H-C-C-N	-1.013	-0.709	0.473
	C-C-N-H	-0.190	-0.417	0.418
	C-C-C-N	2.392	-0.674	0.550
ammonium ion	H-C-N-H	0.000	0.000	0.261
	C-C-N-H	0.000	0.000	0.347
	H-C-C-N	0.000	0.000	0.384
	C-C-C-N	2.732	-0.229	0.485

Table 8. OPLS-AA Fourier Coefficients (kcal/mol) for Torsional Energy Functions[a]

System	Dihedral	V_1	V_2	V_3
peptide[a] φ	C(O)-N-C-C(O)	-2.365	0.912	-0.850
peptide φ'	C(O)-N-C-C	0.000	0.462	0.000
peptide φ"	C(O)-N-C-H	0.000	0.000	0.000
peptide	H-N-Cα-X	0.000	0.000	0.000
peptide ψ	N-C-C(O)-N	1.816	1.222	1.581
peptide ψ'	C-C-C(O)-N	1.173	0.189	-1.200
peptide ψ"	H-C-C(O)-N	0.000	0.000	0.000
peptide	X-Cα-C(O)-O	0.000	0.000	0.000
peptide χ1	N-C-C-C	0.845	-0.962	0.713
peptide χ1	N-C-C-H	0.000	0.000	0.464
peptide χ1	C(O)-C-C-H	0.000	0.000	-0.076
peptide χ1[b]	C(O)-C-C-C	-1.697	-0.456	0.585
χ1, Ser & Thr	N-C-C-O	6.280	-1.467	2.030
Ser & Thr	C(O)-C-C-O	-6.180	0.000	0.000
χ1, Cys	N-C-C-S	0.583	-1.163	0.141
Cys	C(O)-C-C-S	-4.214	-2.114	0.969
5-ethylimidazole	H-C-C-N	0.000	0.000	0.419
	C-C-C-N	2.366	-0.262	0.505
3-ethylindole	H-C-C3-C2	0.000	0.000	-0.480
	H-C-C3-C	0.000	0.000	0.000
	C-C-C3-C2	-0.714	0.000	0.000
	C-C-C3-C	0.000	0.000	0.000
guanidinium ion	H-N-C-N	0.000	3.900	0.000
	C-N-C-N	0.000	7.936	0.000
	C-C-N-C	1.829	0.243	-0.498
	H-C-C-N	0.000	0.000	-0.582
	H-C-N-H	0.000	0.000	0.000

[a] The ϕ and ψ are used in the standard way. The " and ' denote dihedrals which extend to the C_α hydrogen and C_β carbon, respectively. [b] The remaining dihedral parameters for χ1 with the C_α hydrogen are the same as for alkanes, alcohols, and thiols.